NOT CIRCULATE

National Institute on Alcohol Abuse and Alcoholism

RESEARCH

Monograph - 25



Economics and the Prevention of Alcohol-Related Problems

NIAAA
25

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service
National Institutes of Health

Research Monograph No. 25

LIBRARY
ADDICTION RESEARCH CENTER
BALTIMORE, MARYLAND

Economics and the Prevention of Alcohol-Related Problems

Proceedings of a Workshop on Economic and Socioeconomic Issues in the Prevention of Alcohol-Related Problems

October 10–11, 1991
Bethesda, MD

NATIONAL INSTITUTES OF HEALTH
NIH LIBRARY
SEP 26 2007
BLDG 10, 10 CENTER DR
BETHESDA, MD. 20892-1150

Sponsored by:
National Institute on
Alcohol Abuse and Alcoholism

Edited by:
Michael E. Hilton, Ph.D.
Gregory Bloss, M.A.

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service
National Institutes of Health
National Institute on Alcohol Abuse and Alcoholism
5600 Fishers Lane
Rockville, MD 20857

The Editors: Michael E. Hilton, Ph.D., is a health scientist administrator in the Prevention Research Branch of the Division of Clinical and Prevention Research at the National Institute on Alcohol Abuse and Alcoholism (NIAAA). He is responsible for administering the Economics Research Program in the Prevention Research Branch. Gregory Bloss is an economist in the Office of Policy Analysis, NIAAA. He is responsible for surveys and analyses of a range of policy issues pertaining to alcohol-related problems.

All material contained in the monograph, except quoted passages from copyrighted sources, is in the public domain and may be reproduced without permission from the Institute or the authors. Citation of the source is appreciated.

The U.S. Government does not endorse or favor any specific commercial product (or commodity, service, or company). Trade or proprietary names (or company names) appearing in this publication are used only because they are considered essential in the context of the studies reported herein.

The opinions expressed herein are the views of the authors and do not necessarily reflect the official position of the Institute or any other part of the U.S. Department of Health and Human Services.

NIH Publication No. 93–3513
Printed 1993

TABLE OF CONTENTS

Foreword iii

Participants v

Introduction vii

1. "My Kingdom for a Drink . . . ?": A Review of Estimates of the Price Sensitivity of Demand for Alcoholic Beverages 1
   Siu Fai Leung and Charles E. Phelps

2. Taxation of Alcoholic Beverages 33
   Philip J. Cook and Michael J. Moore

3. Alcohol Problems and the Control of Availability: Theoretical and Empirical Issues 59
   Paul J. Gruenewald

4. The Economic Analysis of Addictive Behavior 91
   Michael Grossman

5. Advertising Under the Influence 125
   Henry Saffer

6. Alcohol and the Labor Market 141
   John Mullahy

7. Minimum Drinking Age and Alcohol Availability to Youth: Issues and Research Needs 175
   Alexander C. Wagenaar

8. Measurement Issues in Alcohol Survey Data 201
   Jody L. Sindelar

FOREWORD

It is certainly true that in human life, economic forces are among the many considerations that influence the choices that people make, some of which will be problematic choices. Because of this, such phenomena as economic motivations, consumption choices, prices, taxes, and the influence of advertising are the objects of research interest. Thus, economics takes its place alongside psychology, genetics, biochemistry, and neuroscience as a domain of scientific inquiry that can help us better understand the problems that some people face in connection with drinking.

This volume contains review papers that describe our knowledge of alcohol-related problems from an economic perspective. These papers were presented at a Workshop on Economic and Socioeconomic Issues in the Prevention of Alcohol-Related Problems sponsored by the National Institute on Alcohol Abuse and Alcoholism (NIAAA) in 1991. The reviews presented at that conference should provide useful guidance to those who are trying to advance this line of economic inquiry.

I would like to repeat two remarks that I made at the opening of the workshop. The first is that I would encourage those who think and write about the issues discussed in these pages to have some face-to-face contact with alcoholic patients, to hear from them how they really decide to drink and what they are thinking about when they drink. This can be of tremendous assistance in formulating ideas and refining one's thinking. In this regard, I give economists the same advice that I would give to physicians: There is no substitute for the experience of talking to patients. The best textbooks and the most insightful research reports cannot convey the flavor, the richness, or the intensity that comes out in clinical interviews.

Second, it is crucial to recognize the relationship between the science that NIAAA supports and public policy. There is no escaping that much of what the Institute does is going to have policy implications. Economic research is an area in which these policy implications are readily apparent. In thinking about such research, it is essential to realize that the function of our Institute is to provide the best scientific evidence on the issue at hand—to admit ignorance when we are ignorant and to give the strength of the evidence to the best of our ability when we have the evidence. At NIAAA, whether we are talking about the dopamine receptor or about price elasticities, our policy has been that by adhering to the evidence, we preserve the integrity of our science, and we make our Institute both credible and useful to the people who need its output.

Enoch Gordis, M.D.
Director
National Institute on Alcohol Abuse and Alcoholism

PARTICIPANTS

Gregory Bloss, M.A.
Office of Policy Analysis
National Institute on Alcohol Abuse and Alcoholism
Room 16–95, Parklawn Building
5600 Fishers Lane
Rockville, MD 20857

Frank Chaloupka, Ph.D.
Department of Economics
College of Business Administration
University of Illinois at Chicago
Box 4348
Chicago, IL 60680

Philip Cook, Ph.D.
Institute of Policy Sciences
Duke University
Durham, NC 27706

Dean Gerstein, Ph.D.
National Opinion Research Center
Suite 500
1350 Connecticut Avenue N.W.
Washington, DC 20036

Louis Glicksman, Ph.D.
Program and Services
Evaluation Department
Addiction Research Foundation
University of Western Ontario
London, Ontario N6A 5B9
Canada

Enoch Gordis, M.D.
Director, National Institute on Alcohol Abuse and Alcoholism
Room 16–105, Parklawn Building
5600 Fishers Lane
Rockville, MD 20857

Michael Grossman, Ph.D.
National Bureau of Economic Research, Inc.
269 Mercer Street, 8th floor
New York, NY 10003

Paul Gruenewald, Ph.D.
Prevention Research Center
2532 Durant Avenue
Berkeley, CA 94704

Mike Hilton, Ph.D.
Prevention Research Branch
National Institute on Alcohol Abuse and Alcoholism
Room 13C–23, Parklawn Building
5600 Fishers Lane
Rockville, MD 20857

Ralph Hingson, Ph.D.
Boston University School of Public Health
Social and Behavioral Sciences Department
85 East Newton Street
Boston, MA 02118

Harold Holder, Ph.D.
Prevention Research Center
2532 Durant Avenue
Berkeley, CA 94704

Jan Howard, Ph.D.
Chief, Prevention Research Branch
National Institute on Alcohol Abuse and Alcoholism
Room 13C–23, Parklawn Building
5600 Fishers Lane
Rockville, MD 20857

Siu Fai Leung, Ph.D.
Department of Economics
Harkness Hall 238
University of Rochester
Rochester, NY 14627

Willard Manning, Ph.D.
Division of Health Sciences
Research and Policy
School of Public Health
University of Minnesota
Box 729
420 Delaware Street SE
Minneapolis, MN 55455

Michael Moore, Ph.D.
Fuqua School of Business
Duke University
Durham, NC 27706

John Mullahy, Ph.D.
Department of Economics
Trinity College
300 Summit Street
Hartford, CT 06106

Ethan Nadelmann, Ph.D.
Woodrow Wilson School
Princeton University
Princeton, NJ 08544

Charles Phelps, Ph.D.
Department of Political Science
and Economics
Harkness Hall 306
University of Rochester
Rochester, NY 14627

Henry Saffer, Ph.D.
National Bureau of Economic
Research, Inc.
269 Mercer Street, 8th floor
New York, NY 10003

Jody Sindelar, Ph.D.
Yale University
Department of Epidemiology and
Public Health
P.O. Box 3333
60 College Street
New Haven, CT 06510

Alexander Wagenaar, Ph.D.
Division of Epidemiology
University of Minnesota
1–210 Moos Tower
515 Delaware Street SE
Minneapolis, MN 55455

Kenneth Warner, Ph.D.
Department of Public Health Policy
and Administration
School of Public Health
University of Michigan
1420 Washington Heights
Ann Arbor, MI 48109

Introduction

Michael E. Hilton, Ph.D., and Gregory Bloss, M.A.

This book reviews recent research on a number of topics that we have placed under the general heading, "economic studies of the prevention of alcohol-related problems." This heading is an imperfect description of the book's contents. Not all of the chapters address prevention issues directly, and some of the chapters owe as much to other social science disciplines as they do to economics. Nevertheless, we have adopted this heading because we feel that each chapter provides background information, recent research findings, and theoretical perspectives that will help guide future economic research on prevention issues. As the chapters that follow indicate, research is at a formative stage in many of the areas addressed here. However, the research surveyed in these papers represents vital and challenging areas for further work and development.

We have four primary objectives in publishing this collection of review articles. First, we aim to provide a reasonably comprehensive introduction to the economic literature on the prevention of alcohol-related problems for economists who may be interested in pursuing research in this area. Second, we hope that this book will be useful as a summary of the state of the art in economic analyses of alcohol issues. Third, we wish to acquaint alcohol researchers from other disciplines with the contributions that economic analyses have made to our understanding of the prevention of alcohol-related problems. Finally, we hope that the agenda for future research that emerges in this volume will stimulate work that ultimately helps reduce the prevalence of alcohol-related problems.

Interest in economic aspects of alcoholic beverages has a long history. Taxation of wine and distilled spirits in the United States dates from the 1630's (Aaron and Musto 1981). The "Whiskey Rebellion" of 1794 arose out of perceived inequities in the tax treatment of different alcoholic beverages (Godfrey 1990). The Sub-Committee on Economic Aspects of the Liquor Problem was one of the principal subcommittees established by the turn-of-the-century Committee of Fifty to Investigate the Liquor Problem (Koren 1899; Billings et al. 1905; Levine 1983). Formal economic models of the factors affecting alcohol consumption, alcohol abuse, and alcohol problems are a much more recent phenomenon, with the majority of this research having been conducted in the past decade.

The alcohol field provides rich opportunities for economic researchers to apply their tools. Several issues that arise in studies of alcohol are issues that traditionally have been of concern to economists, including problems in demand modeling and measurement, optimal taxation and the control of externalities, and labor market

issues. There is a need for further research, both theoretical and empirical, in each of these areas. As with economic studies in any particular applied area, research on these generic issues must be guided by the specific characteristics of the topic under study. For example, there is a considerable literature on the demand for beer, wine, and distilled spirits, but the results of these studies provide little information about how to prevent alcohol-related problems. Further, only a fraction of all consumption can be related to problematic outcomes of any kind. Therefore, simple models of the demand for alcoholic beverages need to be supplemented with other theoretical constructs to identify the kinds of prevention policies that confer the greatest benefits at the least cost. Most of the work reviewed in this volume bears on this broad prevention question to some degree, but there is a great deal that remains to be learned.

Despite their relative recency, studies of economic aspects of alcohol consumption have made important contributions to current knowledge about possible prevention strategies. This is perhaps clearest in studies that have examined the effects of changes in prices of alcoholic beverages. As reported in the *Seventh Special Report to the U.S. Congress on Alcohol and Health*:

> Research investigating the relationship between the price of alcoholic beverages and alcohol use problems such as motor vehicle crashes is one of the most promising areas of prevention research (U.S. DHHS 1990, p. 210).

The research alluded to includes several studies that model the effects of changes in alcoholic beverage prices or taxes on traffic crash mortality (Cook 1981; Saffer and Grossman 1987*a,b*), cirrhosis mortality (Cook 1981; Cook and Tauchen 1982), or heavy consumption of alcohol (Coate and Grossman 1988; Grossman et al. 1987). One reason for interest in these findings is that they imply that tax rates, an available policy tool, could be manipulated to reduce the prevalence of alcohol-related problems.

As might be expected, that suggestion is controversial. Indeed, research on the prevention of alcohol-related problems has implications for a range of policy issues, many of which may be sources of contention. A number of the chapters in this book address areas of current or recent policy controversy. As staff at the National Institute on Alcohol Abuse and Alcoholism (NIAAA), we wish to emphasize the distinction between research and advocacy, and to align ourselves clearly with the former. We hope that this book will encourage high-quality research in these areas; it is not intended to promote any particular policy positions. Accordingly, views expressed in this book should be ascribed to the authors of the individual chapters and not to the Institute.

One way that NIAAA stimulates new research is by convening working group conferences. NIAAA held a working group on economic issues in the prevention,

treatment, and epidemiology of alcohol-related problems in September of 1988. That conference led to the development of a program announcement ("Research on Economic and Socioeconomic Issues in the Prevention, Treatment, and Epidemiology of Alcohol Abuse and Alcoholism," December 1988). A variety of investigator-initiated studies were proposed in response to that announcement, and a number of these ultimately were funded.

In anticipation of issuing a new program announcement to supersede the earlier one, NIAAA convened the Working Group on Economic and Socioeconomic Issues in the Prevention of Alcohol-Related Problems in October 1991. The purposes of this conference were to summarize research progress that had been made since the 1988 conference and to identify promising areas for future research. The chapters that follow are based on papers that were prepared for this meeting, with the exception of one chapter that was commissioned after the conference for inclusion in this volume.

Chapter 1, " 'My Kingdom for a Drink . . . ?': A Review of Estimates of the Price Sensitivity of Demand for Alcoholic Beverages" by Siu Fai Leung and Charles Phelps, discusses the price elasticities of alcoholic beverages. This is one of the oldest themes in the literature on the economics of prevention, and it has provided a foundation for much of the other economic work in the field. Drs. Leung and Phelps update the reviews of Ornstein (1980) and Ornstein and Levy (1983), concentrating on demand studies that have appeared since those earlier reviews. A notable trend in this literature is the growing use of survey data in estimating price elasticities. Previously, such studies were usually based on aggregate consumption data. The chapter finds that there is a wide range in published estimates of the price elasticity of demand for alcoholic beverages. The authors look to economic theory for guidance in choosing among alternative estimates, but they conclude that competing estimates must be evaluated on the basis of the quality of the data and the econometric properties of the estimators employed. The chapter closes with an assessment of recent progress in modeling the demand for alcoholic beverages and a discussion of the further progress that can be made by applying more sophisticated econometric techniques.

The subject of price elasticity of demand for alcoholic beverages leads naturally to a discussion of taxation of alcoholic beverages, since taxation is one means of influencing prices. Chapter 2, "Taxation of Alcoholic Beverages" by Philip Cook and Michael Moore, considers issues involved in setting alcohol taxes at appropriate levels. It begins with a brief history of alcohol taxation and then discusses various criteria that may be used to determine the optimal or appropriate levels of tax rates on alcoholic beverages. Consideration is given to alternative notions of fairness, to the public health implications of alcohol taxes, and to the use of alcohol taxes as a means of promoting economic efficiency. The chapter concludes with a discussion of tax equalization across beverage types and an outline of priorities for future research.

In Chapter 3, "Alcohol Problems and the Control of Availability: Theoretical and Empirical Issues," Paul Gruenewald reviews the theoretical and empirical bases of knowledge about alcohol availability and its effects on alcohol consumption and alcohol-related problems. Theoretical approaches considered include those based on: social norms, the distribution of consumption model, consumer demand and the full price of alcoholic beverages, subjective availability, and a theory of routine activities. The empirical literature reviewed includes studies of monopoly versus license systems for the distribution of alcoholic beverages, the geographical density of outlets, hours and days of sale, and laws governing the forms of availability (the principal forms being on-premise and off-premise availability). The author concludes with a discussion of a number of fundamental problems, both theoretical and empirical, that remain to be resolved by future studies of the relationship between alcohol availability and alcohol-related problems.

Chapter 4, "The Economic Analysis of Addictive Behavior" by Michael Grossman, focuses on recent developments in economic models of addictive behavior. The discussion draws on evidence from a range of addictive behaviors, including cigarette smoking, drug use, and gambling, in addition to alcohol consumption. The theoretical framework employed is an extension of consumer demand theory. Consumers are assumed to behave as utility maximizers, even in their decisions to consume harmfully addictive goods. A basic distinction is established between models of "myopic" addictive behavior, in which consumers ignore the effects of current consumption on their future utility in choosing their current consumption level, and models of "rational" addictive behavior, in which consumers' current consumption choices reflect the future effects of that consumption. In either framework, addictive behavior implies that the level of current consumption of an addictive good depends on the levels of previous consumption. The models of addictive behavior discussed in this chapter provide a theoretical basis for the observation that addictive consumption is responsive to price changes. Dr. Grossman surveys the empirical evidence for the rational addiction hypothesis that has accumulated so far and provides some preliminary results from an application of the theory to the consumption of alcoholic beverages. The chapter concludes with a discussion of data on alcohol consumption that may provide further opportunities to test models of addictive behavior.

Economic studies of alcohol advertising are addressed in Chapter 5, "Advertising Under the Influence" by Henry Saffer. Survey and laboratory approaches to studying the effects of advertising on alcohol consumption and problems have long been a familiar part of the prevention literature (see the reviews by Atkin 1987 and Smart 1988). However, economic studies of advertising are a more recent development. The chapter considers five econometric issues that are particularly important in studies of the effects of advertising on the consumption of alcoholic beverages. These five issues provide a framework for the author's review of

econometric studies of the effects of alcohol advertising. The chapter closes with a discussion of promising strategies for future data development and analysis.

In Chapter 6, "Alcohol and the Labor Market," John Mullahy surveys the literature on the relationship between drinking and labor market outcomes. Most of the empirical studies reviewed have focused on such outcome variables as income, earnings, and wages. However, increasing attention also has been paid to such outcomes as labor supply and participation, absenteeism, occupational choice, and the development of human capital. To date, the results of these investigations have often been contradictory. Dr. Mullahy's review identifies and discusses a number of theoretical and econometric issues that might be helpful in reconciling the existing findings. The final section of the chapter considers some promising avenues for future research efforts in this area.

The focus shifts toward youth drinking in Chapter 7, "Minimum Drinking Age and Alcohol Availability to Youth: Issues and Research Needs," by Alexander Wagenaar. This chapter reviews the literature on minimum purchase age laws and how such laws affect alcohol use, traffic crashes, and health and social problems. Much of the prevention literature on youth drinking has focused on determining the impact of changes in the minimum purchase age laws. With the nationwide adoption of a purchase age of 21, and with a reasonable consensus that this change has reduced alcohol-related problems among youth, research attention now seems to be pointed toward other aspects of youth drinking, such as sources of access to alcohol, the settings in which underage drinkers typically consume alcohol, and the enforcement practices of State alcoholic beverage control agencies regarding underage drinking. Dr. Wagenaar contends that research into these issues would benefit from an eclectic methodological approach that includes time-series analyses of natural experiments, controlled community trials, purchase attempt studies, and observational studies of youth drinking occasions.

In the final chapter, Jody Sindelar discusses "Measurement Issues in Alcohol Survey Data." This chapter introduces some of the data and measurement problems that economists will encounter in using survey data on alcohol consumption and alcohol-related problems. Many of these problems are well known to researchers in the alcohol field but will be unfamiliar to other researchers. The chapter provides an empirical example that illustrates how the choice of a particular measure of alcohol consumption or alcohol problems can affect the outcome of an analysis. This exercise should prompt economists to be alert to the potential consequences that measurement choices might have. The chapter concludes with a discussion of the availability (or lack thereof) of data that are of particular interest to economists seeking to do research in the alcohol area.

This volume is the work of many people who deserve grateful acknowledgment here. We are, of course, especially appreciative of the efforts of the authors for the chapters they contributed to this volume. They in turn benefited from the con-

structive criticisms of discussants at the 1991 conference. The names of all conference participants appear on pages v–vi. Jan Howard, chief of the Prevention Research Branch at NIAAA, made the initial decision to hold this working group conference and should be credited with the inception of this project. Susan Farrell, former director of the Office of Policy Analysis at NIAAA, gave substantial assistance to the project both by providing advice and by committing the resources of her office. Nancy Colladay's skill in planning and managing the logistics of the workshop conference were invaluable.

REFERENCES

Aaron, P., and Musto, D. Temperance and prohibition in America: A historical overview. In: Moore, M.H., and Gerstein, D.R., eds. *Alcohol and Public Policy: Beyond the Shadow of Prohibition.* Washington: National Academy Press, 1981. pp. 127–181.

Atkin, C. Alcoholic-beverage advertising: Its content and impact. In: Holder, H., and Mello, N., eds. *Control Issues in Alcohol Abuse Prevention: Strategies for States and Communities.* Suppl. 1. Greenwich, CT: JAI Press, 1987. pp. 267–287.

Billings, J.S.; Elliot, C.W.; Farnham, H.W.; Greene, J.L.; and Peabody, F.G. *The Liquor Problem: A Summary of Investigations Conducted by the Committee of Fifty, 1893–1903.* Boston: Houghton Mifflin Company, 1905.

Coate, D., and Grossman, M. Effects of alcoholic beverage prices and legal ages on youth alcohol use. *J Law Econ* 31(1):145–171, 1988.

Cook, P.J. The effect of liquor taxes on drinking, cirrhosis, and auto accidents. In: Moore, M.H., and Gerstein, D.R., eds. *Alcohol and Public Policy: Beyond the Shadow of Prohibition.* Washington: National Academy Press, 1981. pp. 255–285.

Cook, P.J., and Tauchen, G. The effect of liquor taxes on heavy drinking. *Bell J Econ* 13(2):379–390, 1982.

Godfrey, J. Alcohol taxes: From the Whiskey Rebellion to the 1990 budget summit. *Tax Notes* July 9, 1990, pp. 137–140.

Grossman, M.; Coate, D.; and Arluck, G.M. Price sensitivity and alcoholic beverages in the United States: Youth alcohol consumption. In: Holder, H., ed. *Control Issues in Alcohol Abuse Prevention: Strategies for States and Communities.* Greenwich, CT: JAI Press, 1987. pp. 169–198.

Koren, J. *Economic Aspects of the Liquor Problem: An Investigation Made Under the Direction of a Sub-Committee of the Committee of Fifty.* Boston: Houghton Mifflin Company, 1899.

Levine, H.G. The committee of fifty and the origins of alcohol control. *J Drug Issues* 13(1):95–116, 1983.

Ornstein, S. Control of alcohol consumption through price increases. *J Stud Alcohol* 41(9):807–818, 1980.

Ornstein, S.I., and Levy, D. Price and income elasticities and the demand for alcoholic beverages. In: Galanter, M., ed. *Recent Developments in Alcoholism.* Vol. 1. New York: Plenum, 1983. pp. 303–345.

Saffer, H., and Grossman, M. Beer taxes, the legal drinking age, and youth motor vehicle fatalities. *J Legal Stud* 16(2):351–374, 1987*a*.

Saffer, H., and Grossman, M. Drinking age laws and highway mortality rates: Cause and effect. *Econ Inquiry* 25(3):403–417, 1987*b*.

Smart, R.D. Does alcohol advertising affect overall consumption? A review of empirical studies. *J Stud Alcohol* 49(4):314–323, 1988.

U.S. Department of Health and Human Services. *Seventh Special Report to the U.S. Congress on Alcohol and Health.* DHHS Pub. No. (ADM) 90–1656. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1990.

"My Kingdom for a Drink. . . ?": A Review of Estimates of the Price Sensitivity of Demand for Alcoholic Beverages

Siu Fai Leung, Ph.D., and Charles E. Phelps, Ph.D.

INTRODUCTION

A main focus of economic studies of alcohol is to examine the price sensitivity of the demand for alcoholic beverages. Similar to other empirical demand analyses, most of the studies center around the estimation of elasticities.[1] A good deal of econometric work has appeared since Stone's first attempt to estimate the price elasticities of demand for beer and spirits in 1945 (Stone 1945).

The most extensive surveys of the literature are those by Ornstein (1980) and Ornstein and Levy (1983).[2] They found that the estimates of the price elasticity of demand range from –1.39 to +0.01 for beer, from –3.65 to +0.02 for distilled spirits, and from –3.28 to –0.06 for wine. The estimates of the cross-price elasticity of demand range from –1.01 (spirits price on wine consumption) to +2.55 (beer price on wine consumption).[3] These findings on the whole vindicate the economic theory that demand curves for alcoholic beverages are negatively sloped. Nevertheless, given such a wide range of estimates, Ornstein (1980, p. 817) concluded his survey with a rather negative assessment of the literature:

> The main lesson to be learned from this survey is that demand estimation is a difficult and inexact business. The results are strongly affected by the data and econometric techniques used. One should be cautious before adopting the results of any given study, and especially cautious if such results are to be used as a basis for public policy.

---

[1] *Elasticity, a concept created by Alfred Marshall, plays a fundamental role in economics (especially in demand analysis). See Keynes (1933) and Newman (1987) for historical accounts of the concept. Deaton and Muellbauer (1980b) offer several reasons to explain why the notion of elasticity has been so widely accepted and used since its inception.*

[2] *The two literature surveys cover the same materials, and the viewpoints are basically the same, except that the one by Ornstein and Levy provides a much more elaborate discussion of the literature.*

[3] *The price (own-price) elasticity of demand is defined by* $(P_i/Q_i)(dQ_i/dP_i)$, *where* P_i *and* Q_i *denote the price and quantity demanded of good i. The cross-price elasticity of demand of good i with respect to the price of good j is defined by* $(P_j/Q_i)(dQ_i/dP_j)$.

Despite the lack of consensus among the findings, Ornstein and Levy attempted to narrow the range of estimates. After a detailed comparison and evaluation of the studies, they reported that their "summary estimates of own-price elasticities for beer, wine, and distilled spirits were –0.3, –1.0, and –1.5, respectively. These are crude at best, particularly for wine, but seem the best available" (Ornstein and Levy 1983, p. 343).[4] No summary estimates were given for cross-price elasticities.

All the econometric studies reviewed in the Ornstein and Ornstein and Levy surveys were based on aggregate data and were done more than 10 years ago.[5] In the past decade, research on the price sensitivity of alcohol consumption has continued, and a major development has emerged. The proliferation of economic work based on individual data (microdata) and the development of econometric techniques for these data (microeconometrics) have begun to exert their influence on alcohol research.[6] In contrast to previous work on the demand for alcoholic beverages, which employs exclusively aggregate data, several studies based on individual data have recently appeared (e.g., Grossman et al. 1987; Coate and Grossman 1988; Heien and Pompelli 1989; Atkinson et al. 1990).

The availability and use of individual data can shed light on debates that cannot be resolved using aggregate data. One good example is the debate on whether heavy drinkers are responsive to price changes. Price elasticities estimated from aggregate data measure only the effect of price on the alcohol consumption of the population as a whole. From a public health policy point of view, Room (1984) contended that such a broad measure is not of central interest because what really matters is the effect of price on high-risk drinking patterns.

The alcohol literature frequently asserts that price will only affect moderate and trouble-free drinkers (Room 1984) and that increases in prices on alcoholic beverages will have no effect on heavy drinkers and problem drinkers because they are either addicted to or physically dependent on alcohol (Pittman 1980). Therefore, increases in taxes on alcoholic beverages will only affect moderate and trouble-free drinkers. To evaluate these claims, individual data are needed because there is no direct way to gauge from aggregate data the impact of price increases on the alcohol consumption of heavy drinkers and problem drinkers.[7] Consequently, previous research on aggregate price elasticities of

---

[4] *To facilitate comparisons with other estimates in the literature, we will call these three values of price elasticities the "Ornstein and Levy benchmark."*

[5] *Different forms of aggregate data have been used: country-wide time series and cross-sections, State cross-sections in the United States, and province cross-sections in Canada.*

[6] *See Amemiya (1985), Greene (1990), Maddala (1983), and Pudney (1989) for surveys of the literature.*

[7] *By using some proxies for heavy drinking and problem drinking, aggregate data can nevertheless be used to offer a partial solution to the problem. For instance, there are studies which employ aggregate data to relate the adverse effects of heavy drinking and problem drinking, such as liver cirrhosis mortality rates and automobile accidents, to the prices of alcoholic beverages (Seeley 1960; Cook 1981).*

alcohol has not yet provided complete and direct support for the use of price as a public health policy instrument.[8] However (see footnote 7), some important indirect evidence is available on this issue.

In this paper, we will critically review the literature on the price sensitivity of alcoholic beverages that has appeared since the Ornstein and Ornstein and Levy surveys. The literature can now be divided into two broad categories: aggregate data studies and individual data studies. Such a distinction is useful because each category of work addresses different issues and employs different estimation methods. As expected, the findings vary, and there is no consensus on which estimates are better. We will first investigate whether economic theory can produce any practical criteria to select estimates. In particular, we will examine the two most popular theories in the literature: the revenue maximization model advocated by Ornstein (1980) and Ornstein and Levy (1983), and the welfare maximization model developed by Pogue and Sgontz (1989).

We will then summarize and critically review the objectives and findings of the alcohol demand studies that have appeared since the Ornstein and Levy survey.[9] A special emphasis will be put on those studies that examine the demand for alcoholic beverages in the United States. Following these individual reviews, we will offer an overview of the literature and an agenda for future research. We end with our concluding remarks.

ECONOMIC THEORY

Although the empirical evidence presents a rather wide range of estimates of the price elasticity of demand for alcoholic beverages, a more fundamental issue remains as yet unspecified: Why do we wish to know the demand elasticity? An obvious response is that good estimates of the price responsiveness are necessary to establish public policy regarding alcoholic beverages. To do this, a model of government objectives is needed. In addition, if we can assume some model of government behavior, then we can infer something about the relative elasticities of demand from the various government experiments with different levels of taxation. Two models of government behavior are suggested in the literature, and unfortunately, they lead to different choices of tax levels. Thus, unless we can assume or demonstrate one particular government objective function, we may not be able to make much use of this strategy. The two competing models of government behavior are revenue maximization and welfare maximization. We discuss each in turn next.

---

[8] *The premise that alcohol policy matters only to the extent that it alters heavy drinking also deserves rethinking. While it is true that much of the social cost of alcohol arises from heavy drinking (including liver cirrhosis, fatal vehicle crashes, and domestic violence), social consequences arise from more moderate drinking as well, including some contribution to vehicle fatalities (Phelps 1988) and personal health problems.*

[9] *Although we have made an extensive search of the literature to locate all the relevant alcohol demand studies published since the 1983 Ornstein and Levy survey, we are aware that omissions are unavoidable.*

Revenue Maximization

One model of government behavior assumes a goal of revenue maximization. This approach has widely been adopted [see, for example, Ornstein (1980), Ornstein and Levy (1983), Duffy (1983), Walsh (1983), and Levy and Sheflin (1985)]. Ornstein (1980, p. 809) stated, "To the extent governments are predominantly interested in maximizing revenues, it is unlikely they would choose to tax high price-elastic goods, since tax increases on such goods would tend to reduce total tax revenues." The argument is based on Geary's (1973) formula on the relationship between the elasticity of tax revenue with respect to tax rate and the price elasticity of demand:

$$e_t = 1 + [t/(p + t)]\, e_p \qquad (1)$$

where $e_t = (t/R)(dR/dt)$, $e_p = (p^*/q)(dq/dp^*)$, t = per unit tax rate, p = price net of taxes, $p^* = p + t$, q = quantity, and R = tax revenue = tq.[10] Ornstein (1980) deduced from equation 1 that e_t will be positive whenever e_p is unit elastic or inelastic and that the higher the tax as a proportion of selling price, the lower must be the price elasticity. He concluded that given the high rates of taxation on alcoholic beverages, the demand may well be inelastic. More specifically, he predicted that the absolute values of the price elasticities of distilled spirits, wine, and beer should be in increasing order because the tax rate is highest for distilled spirits, followed by that for wine and then by that for beer.

There are several fundamental problems with Ornstein's theory. First, $e_p > -1$ is only a sufficient condition for e_t to be positive. It is clearly not a necessary condition since e_t can be positive even if e_p is elastic (< -1). Second, contrary to Ornstein's contention, a revenue-maximizing government would set the tax rate such that the equilibrium consumption occurs in the elastic region of the demand curve. This follows from the fact that $dR/dt = 0$ implies $e_t = 0$; therefore, $e_p = -(p + t)/t < -1$, i.e., elastic demand. Third, despite the widespread adoption of equation 1 in the alcohol literature, the equation is derived from a very restrictive and questionable assumption. A more general derivation shows that the correct formula is given by

$$e_t = 1 + [t/(p + t)] \div \{1/e_p^d - [p/(p+ t)][1/e_p^s]\} \qquad (2)$$

where e_p^d and e_p^s are the demand and supply elasticities.[11] Only if $e_p^s = \infty$, will equation 2 be identical to equation 1. Hence, equation 1 implicitly assumes a horizontal supply curve. Nevertheless, the literature has provided neither the theory

[10] *To derive equation 1, differentiate R with respect to t: $dR/dt = q + tdq/dt = q + (tq/p^*)(p^*/q)(dq/dp^*)/(dp^*/dt) = q[1 + (t/p^*)(p^*/q)(dq/dp^*)]$. Note that demand q is a function of p^*.*

[11] *To see this, let the supply and demand curves be denoted by S(q) and D(q). Supply equals demand at equilibrium; therefore, $t + p = p^*$ implies that $t + S(q) = D(q)$. Hence $dq/dt = 1/(dD/dq - dS/dq) = 1/(dp^*/dq - dp/dq) = 1/[1/(dq/dp^*) - 1/(dq/dp)] = 1/[(p^*/q)(1/e_p^d) - (p/q)(1/e_p^s)] = (q/p^*)\{1/[1/e_p^d - (p/p^*)(1/e_p^s)]\}$, and equation 2 follows.*

nor the evidence to show that the supply curves for alcoholic beverages are perfectly elastic; therefore, the use of equation 1 is unwarranted.[12] Equation 2 shows that there is no particular restriction on the magnitude of e_p^d, even for a revenue-maximizing government. It is therefore not surprising to find that Ornstein's prediction on the ranking of the demand elasticities of distilled spirits, wine, and beer is not supported by empirical findings in the literature, even if the presumption of revenue maximization is correct.

Welfare Maximization

An alternative model of government behavior says that governments do (or should) set taxes to maximize social welfare. One formal model of this approach was set forth by Pogue and Sgontz (1989). In this model, one taxes alcoholic beverages to control social costs, and the tax is chosen to minimize the sum of welfare costs to those taxed (drinkers) and those harmed by the drinkers (e.g., those killed in vehicle crashes, victims of domestic violence precipitated by alcohol consumption). Pogue and Sgontz developed a model with two types of drinkers (abusive and nonabusive), but they did not allow for different types of alcoholic beverages or different tax levels for them. Nevertheless, it is clear that the optimal tax in their model depends on (among other things) the demand elasticities. It also focuses attention on the importance of the *relative* price responsiveness for abusers (who are assumed to create social costs) and nonabusers (who are assumed to create no social costs). Holding other things equal (including the price responsiveness of abusers), the optimal tax in their model increases with the price responsiveness of abusers. By contrast, the optimal tax *decreases* with the price responsiveness of nonabusers, holding other things constant (including the price responsiveness of nonabusers).

In a separate study in the welfare-maximizing tradition, Phelps (1988) did not assume separate elasticities for abusers and nonabusers (a key feature of the model of Pogue and Sgontz), but rather assumed a single elasticity summarizing the demands by all drinkers. In that study, the welfare losses to drinkers from a given tax fall as the demand elasticity falls to zero, but the efficacy of higher taxes in reducing social costs also falls. Thus, the optimal tax in such a model does not necessarily move inversely to the demand elasticity. This approach thus

[12] *A horizontal supply curve implies that the alcoholic beverage market is perfectly competitive. Based on the dominance (in terms of domestic market share) of the top two brewers (Anheuser-Busch, 45.9 percent; and Miller, 23.1 percent), Cook and Moore (chapter 2) reject the notion that the U.S. beer industry is perfectly competitive. Instead, they argue that the beer market is oligopolistic. While their argument is plausible, more research on the market structure of the alcohol industry is needed because dominance of market share alone is a crude measure of the lack of perfect competition. The relationship between market concentration and market structure is a long-debated subject in industrial organization. In his seminal work "Why Regulate Utilities?" Demsetz (1968, pp. 59–60) argued (and illustrated) that "we have no theory that allows us to deduce from the observable degree of concentration in a particular market whether or not price and output are competitive."*

offers a prescription for government behavior that may conflict with the revenue-maximizing model.

Because of potential conflicts in the prescriptions of various models of government behavior, we cannot draw inferences about the validity of demand estimates without having strong reason to believe one or another model of government behavior. In such a case, the only criteria to judge estimates of the demand for alcoholic beverages are the intrinsic econometric properties of the estimates. With this in mind, we turn now to a review of studies of demand that have appeared recently in the literature.

REVIEW OF THE LITERATURE: STUDIES BASED ON AGGREGATE DATA

Cook (1981)

In a widely cited study, Cook (1981) replicated Simon's (1966) nonparametric quasi-experimental study by using State tax changes that occurred between 1960 and 1975 in the United States.[13] With some minor modifications of Simon's procedure, he found that the median of the price elasticities of liquor is –1.6. He then used the same methodology to estimate the price elasticities of the auto fatality rate and liver cirrhosis death rate. He found that the medians of the price elasticities are –0.7 for auto fatalities and –0.9 for cirrhosis deaths.

Cook maintained that one of the great virtues of Simon's quasi-experimental approach is that the use of States that did not change taxes to control for nonprice influences on liquor is probably more reliable than the econometric technique of specifying one or more control variables in the regression, because these influences are not likely to be captured by any short list of control variables. While Simon's approach provides an alternative to the regression approach, more work is needed before it can be used with confidence. A potential problem with the quasi-experimental method is that the statistical properties of the estimates have not yet been established. As such, the use of the terminologies "biased" and "unbiased" in Cook (1981) is premature.

The significance tests performed in Simon and Cook are also questionable. For example, the nonparametric sign test requires that the observations are independent. However, some States in their samples contribute more than one observation, so it is unclear how the independence assumption can be justified. For instance, unless there are drastic changes in laws or population compositions, a State that has a highly inelastic demand for liquor in one year will tend to have a highly inelastic demand for liquor in another year. Furthermore, the use of control States in Simon's approach is based on the presumption that unmeasured factors that influence consumption in one State in any given year will also be influential in other States in that year, which suggests that the changes in liquor consumption among the States cannot be independent.

[13] *The sample period used by Simon (1966) is 1950–61.*

Cook and Tauchen (1982)

Since Simon's nonparametric technique "does not generate a useable estimate of the magnitude of the effect in question" (Cook 1982), Cook and Tauchen (1982) employed a parametric estimation technique to a U.S. time-series cross-sectional data set from 30 States for the period 1962–77. The dependent variables examined are liquor consumption and age-standardized cirrhosis mortality. The model is

$$\log(liq_{it}) = \alpha_0 Tax_{i,t} + \alpha_1 Tax_{i,t-1} + \theta_0 Inc_{i,t} + \theta_1 Inc_{i,t-1} + \nu_i + \gamma_t + e_{it}$$

where liq_{it} = per capita liquor consumption in State i at year t, $Tax_{i,t}$ = liquor tax rate in State i at year t, $Inc_{i,t}$ = per capita income in State i at year t, ν_i = time-invariant factors specific to State i, γ_t = time-varying factors common to all States, and e_{it} = random effect (time-varying factors specific to State i). The tax and income variables are lagged once because the authors found that longer lags are rejected. The estimation procedure is two-step generalized least-squares: The first step estimates the first-order autocorrelation parameter of the random effect, and the second one is a standard generalized least-squares procedure. They found that the price and income elasticities of liquor at the 1968 mean price and income are –1.8 and 0.43, respectively. These estimates are clearly in line with Ornstein and Levy's (1983) benchmark.

Since Cook and Tauchen's model contains only two explanatory variables (income and tax), it is not clear whether the two fixed effects ν_i and γ_t can capture all the other relevant factors that might affect liquor consumption and cirrhosis mortality. For example, the tax rates (and prices) of beer and wine, as well as drinking laws and regulations (e.g., minimum legal drinking age), vary over time and across States. It is highly unlikely that these factors can be captured by the fixed effects ν_i and γ_t. Despite these problems, Cook and Tauchen provided an important contribution to the literature because, unlike most other alcohol demand studies which employ either time-series or cross-sectional data, they exploited both types of data and used the dummy variable approach to estimate the model. They also evaluated their assumptions with careful diagnostic procedures. It would be useful to apply their methodology to study the demand for beer and wine.[14]

Schweitzer, Intriligator, and Salehi (1983)

The purpose of Schweitzer et al. (1983) was "to specify, to estimate and to use an econometric model of alcoholism which

[14] *In an attempt to estimate a pooled time-series cross-sectional model for spirits and beer (with a State-specific fixed effect but no time-specific fixed effect), Ornstein and Hanssens (1985) could not obtain unique estimates of the State dummies, apparently because of serious multicollinearity between the State dummies and the regulatory dummy variables. As a result, they abandoned the dummy variable approach and just ran OLS on the pooled data. Whether this is a generic problem of the dummy variable approach deserves further attention.*

would incorporate its causes, its effects, and its possible control." The data set consists of a cross-section of 35 States in the United States in 1975. They considered the following four-equation simultaneous equations model:

$$log(CB) = a_1 log(ASM) + X_1 a + u_1$$

$$log(CS) = b_1 log(ASM) + X_2 b + u_2$$

$$log(ASM) = c_1 log(CB) + c_2 log(CS) + X_3 c + u_3$$

$$log(ARM) = d_1 log(ASM) + d_2 log(CB) + d_3 log(CS) + X_4 d + u_4$$

where CB = consumption of beer, CS = consumption of spirits, ASM = alcoholism, ARM = alcohol-related mortality, and X_i (i = 1,..,4) is a vector of explanatory variables that include per capita income, urbanization, religion, tourism, alcoholism rehabilitation expenditures, and unemployment rate. Only beer and distilled spirits are examined, and the estimates of the own-price and cross-price elasticities for the reduced and structural forms are contained in table 1.

The results do not accord with the Ornstein and Levy benchmark in that spirits have a more inelastic demand than beer in the reduced-form model. On the other hand, the structural estimates do not seem to be credible because the price elasticity of spirits is positive. Notice however, that none of the estimates in table 1 is statistically significant.[15] The Schweitzer et al. study suggests that price control policies are not as effective as other variables such as minimum age laws and unemployment rate. Although they emphasized throughout their study that a simultaneous equations approach is more useful and powerful, the results really do not support their claim because none of the estimates of the structural parameters a_1, b_1, c_1, c_2, d_1, d_2, and d_3 turn out to be significant.

TABLE 1

Schweitzer, Intriligator, and Salehi estimates of price elasticities

	Beer	Spirits
Beer	–0.27[a] –0.47*[b]	0.33[a] –0.11[b]
Spirits	0.16[a] 0.33[b]	–0.13[a] 0.35[b]

Notes: Each number in the table represents the elasticity of the column variable with respect to the price of the row variable.
[a] Reduced form model.
[b] Structural model.
** Statistically significant at the 10-percent level (two-tailed test).*

Levy and Sheflin (1983, 1985)

Levy and Sheflin maintained that attempts to determine the cross-price effects in studies of demand for beer, distilled spirits, and wine have been unsuccessful because the estimates of cross-price elasticities are generally inconsistent across, and even within, studies, and because some studies find the implausible result

[15] *The price elasticity of beer in the structural model is marginally significant at the 10-percent level. In fact, most of the estimates of the parameters of the reduced and the structural forms are insignificant.*

that some alcoholic beverage categories are complements (instead of substitutes).[16] The failure to produce satisfactory estimates of cross-price effects makes it difficult to gauge the effect of tax-induced price changes on total alcohol consumption; therefore, the current literature provides a questionable basis for public policy analysis.

In view of these problems, Levy and Sheflin advocated an alternative strategy of estimating a single demand equation for all alcoholic beverages. This approach would subsume cross-price effects and produce estimates of price elasticity net of substitution possibilities among beverages. As long as the relative prices of the beverages remain constant, the estimates will provide a sound basis for public policy analysis on total alcohol consumption. They estimated the model

$$dlog(C) = a + bdlog(P) + cdlog(I) + u$$

where C = per capita consumption of alcoholic beverages, P = relative price of alcoholic beverages, I = real per capita disposable personal income, and the d's indicate first differences (to remove time trend and reduce serial correlation). The estimation employs annual U.S. time-series data over the period 1940–80.

Two measures of the consumption variable C are investigated, and a separate regression is run for each measure. The first one measures C by volume and is constructed by multiplying annual data on the apparent consumption of wine and distilled spirits and tax-paid withdrawals of beer by their respective average alcohol contents, and cumulating to obtain the annual volume of pure alcohol consumed. The second one measures C by expenditures and is obtained from dividing personal consumption expenditures on alcoholic beverages by the price of alcoholic beverages. The two regressions yield similar estimates: –0.5 for the price elasticity and 0.4 for the income elasticity. All the elasticity estimates are statistically significant at the 5-percent level.

There are three problems in the analysis. First, the price variable P for the first regression is obtained by dividing nominal expenditures on alcoholic beverages by the volume of alcohol consumed. Thus, the dependent variable (volume) appears on both the right-hand side and the left-hand side of the regression. The assumption that P is uncorrelated with the error term is then violated, and the estimates will no longer be unbiased and consistent. Second, the only other explanatory variable included in the regressions is the proportion of population aged 16–20. It is likely that there are many missing variables that could affect the validity of the estimates. Third, as Nelson (1990, p. 85) has pointed out, aggregating all three types of alcoholic

[16] *Note that some forms of drinking directly create complementarity, such as the combination of a distilled spirit (gin or vodka) and a wine (vermouth) in the creation of martinis, or the consumption of "boilermakers," i.e., a shot of whiskey followed by a glass of beer. While these may appear whimsical, one should not rule out complementarity on the basis of theory. Binge drinkers, for example, often consume multiple drinks of different types, including wine, beer, and distilled spirits.*

beverages into a single alcohol consumption variable presupposes that "people consume a beverage solely for its absolute alcohol content and that the ethanol content of each beverage is similar across brands and states." Such an assumption is clearly questionable.

Ornstein and Hanssens (1985)

The objective of the Ornstein and Hanssens (1985) study was to assess the effectiveness of various alcohol control laws on the consumption of alcohol. The data set is a pooled time-series cross-sectional sample of 50 States and the District of Columbia for the period 1974–78 for distilled spirits and of 48 States and the District of Columbia for the period 1976–78 for beer. Wine is excluded from the analysis because of a lack of price data. There are 26 explanatory variables grouped in 5 categories: economic variables, sociodemographic variables, regulations on retail availability, price controls, advertising controls, and a dummy variable for monopoly States versus license States. Although the data are pooled time-series cross-section, they employ ordinary least squares (OLS) to estimate a standard semilogarithmic model because neither the fixed effect model nor the variance components model that they attempted yielded satisfactory results.

Ornstein and Hanssens found that the price elasticity of beer is –0.12 for the whole sample and –0.17 for the license States. For spirits, the price elasticities are –0.92 for the whole sample and –0.66 for the license States.[17] These estimates are generally significant at either the 5-percent or 10-percent level. Hence, the demands for distilled spirits and for beer are inelastic. In comparison with the other estimates, Ornstein and Hanssens found that control laws affecting price have the greatest impact on the consumption of distilled spirits. For beer, the small price elasticity suggests that control laws affecting price have little effect on the consumption of beer. The most effective control laws are minimum legal age and Sunday sales. We will return to Ornstein and Hanssens' findings in our discussion of Nelson (1990).

Uri (1986)

In an attempt to measure the extent of substitution among seven beverages (soft drinks, beer, fruit drinks, drink mixes, spirits, wine, and bottled water), Uri (1986) employed a standard double logarithmic regression model for each type of beverage:

$$log(q_i) = a_{i0} + \sum_{j=1}^{m} a_{ij} log\,(p_j) + b_i log(y) + \sum_{j=1}^{n} c_{ij} log\,(x_j) + u_i$$

where q_i = quantity of beverage i demanded, p_i = price of beverage i, y = income, and x_j = other explanatory variables (e.g., the proportions of population in various age groups, temperature, and death rate due to liver cirrhosis). He used cross-sec-

[17] *Since there are several specifications for each regression and the estimates of the price elasticities are very close to each other, the numbers reported here are averages of the estimates found in Ornstein and Hanssens (1985).*

tional data (50 States plus the District of Columbia) for the year 1982, and the method of estimation was maximum likelihood. Since his likelihood ratio test cannot reject the symmetry hypothesis ($a_{ij} = a_{ji}$), he imposed the restrictions on the model. In general, he found that changing relative prices do have statistically significant effects on beverage consumption. Table 2 contains some of his estimates.

The results are rather unusual: Beer is demand elastic while wine is demand inelastic. As in many other studies, the price data used here do not seem to have high quality because the price is obtained from dividing total retail dollars by total quantity sold. Some preliminary evidence of this deficiency is reported in Nelson (1990). In contrast with some of the findings in the literature, the table indicates clearly that beer, spirits, and wine are all substitutes. We will return to this problem in the section entitled General Discussion.

TABLE 2

Uri estimates of price elasticities

	Beer	Spirits	Wine
Beer	−1.07†	0.31†	0.48*
Spirits	0.31†	−1.21†	0.62†
Wine	0.48*	0.62†	−0.88†

Notes: Each number in the table represents the elasticity of the column variable with respect to the price of the row variable.
** Statistically significant at the 10-percent level (two-tailed test).*
† Statistically significant at the 5-percent level (two-tailed test).

Nelson (1990)

Similar to that of Ornstein and Hanssens (1985), the main theme of Nelson (1990) was to examine the effect of State monopolies and other regulations on the consumption of alcoholic beverages. It differs from previous studies in several respects. First, unlike earlier work that only examined beer and spirits consumption, Nelson manages to include wine consumption in the analysis. Second, the quality of the price data is better than that of others. Third, he applied Leamer and Leonard's (1983) Bayesian extreme bound analysis to evaluate the robustness of the OLS estimates. The models are standard semilogarithmic demand equations, one for each type of alcoholic beverage. Each equation contains only the price of one other type of alcoholic beverage (the price of spirits in the beer equation, and the price of beer in the wine and spirits equations). The data are U.S. State cross-sectional for the year 1982.

As shown in table 3, Nelson found that the own-price elasticities are all significant and negative. Beer demand is inelastic. For each regression specification, wine has a more elastic demand than spirits, which is inconsistent with the Ornstein and Levy benchmark. The cross-price effects are all insignificant. As for the main objective of the paper, Nelson found that State monopolies do not have any effect on the consumption of all three types of alcoholic beverages. This contrasts with Ornstein and Hanssens (1985), who found that monopoly States have a significantly lower level of spirits consumption than license States. He also found that minimum drinking age laws

TABLE 3

Nelson estimates of price elasticities

	Beer	Spirits	Wine
Beer	–0.364†[a] –0.561†[b]	0.160[b]	–0.603[b]
Spirits	0.089[b]	–1.149†[a] –0.890†[b]	NA
Wine	NA	NA	–2.291†[a] –1.862†[b]

Notes: Each number in the table represents the elasticity of the column variable with respect to the price of the row variable.
[a] The model contains six focus regressors.
[b] The model contains six more regressors in addition to the six focus variables.
** Statistically significant at the 10-percent level (two-tailed test).*
† Statistically significant at the 5-percent level (two-tailed test).
NA = Not available because the price variable is not included in the regression.

have no effect on beer consumption, which is again different from Ornstein and Hanssens' results. Further research is needed to resolve these discrepancies.

Foreign Demand Studies

Apart from using different data sets, foreign alcohol demand studies also tend to adopt a different approach. While domestic studies utilize exclusively the single-equation methodology to estimate the demand for alcoholic beverages, a significant portion of foreign work employs a system (i.e., system-wide) approach. The system approach emphasizes the virtue of modeling the demand for goods in a complete system of equations. It is superior to the single-equation approach because it can exploit the restrictions provided by economic theory and allow the demands for different commodities to interact.

Two particular system models have been applied to investigate alcohol demand: the Rotterdam system by Clements and Johnson (1983) and Selvanathan (1988) and the almost ideal demand system (AIDS) by Thom (1984) and Jones (1989). We will briefly describe the two models here.[18]

The Rotterdam system was developed by Theil and Barten in Rotterdam around 1965 (see Deaton and Muellbauer 1980*b*). The model focuses on the equation

$$w_{it} dlog(q_{it}) = \gamma_i dlog(m_t^*) + \sum_{j=1}^{n} \beta_{ij} dlog(p_{jt})$$

where $i = 1, 2, ..., n$ and $t = 1, 2, ..., T$ (n goods and T periods), d denotes total differential, q_{it} = quantity of good i consumed in period t, w_{it} = ratio of expenditure on good i in period t to total expenditure in period t (i.e., budget share of good i in period t), p_{ij} = price of good j in period t, and $dlog(m_t^*) = dlog(m_t) - \Sigma_{i=1}^{n} w_{it} dlog(p_{it})$, where $m_t = \Sigma_{i=1}^{n} p_{it} q_{it}$ (total expenditure in period t). In empirical work, the total differential $dlog(z_t)$ is approximated by $log(z_t) - log(z_{t-1})$.

Demand theory implies two restrictions on the parameters β_{ij}. The homogeneity restriction states that $\Sigma_{j=1}^{n} \beta_{ij} = 0$ for all i and the symmetry restriction requires that $\beta_{ij} = \beta_{ji}$ for all $i \neq j$.

The homogeneity restriction follows from the property that the Hicksian

[18] *Deaton and Muellbauer (1980b) and Deaton (1986) contain excellent discussions of these and other system models.*

demand function is homogeneous of degree 0. A proportionate change in the prices of all the n goods would (should) not affect the Hicksian demand for any of the n goods. Thus, if this theory is correct, then $\Sigma_{j=1}^{n} \beta_{ij} = 0$ must be true for every i. Demand theory also implies that, holding real income constant, the effect of an increase in the price of good j on the demand for good i is equal to the effect of an increase in the price of good i on the demand for good j. If this symmetry theory is correct, then $\beta_{ij} = \beta_{ji}$ for all $i \neq j$.[19]

Clements and Johnson (1983) estimated a slightly more complicated version of the model. Because they assumed that the utility function is separable in alcoholic beverages and all other goods, they could focus on the three beverages and ignore all other goods. In particular, they considered a conditional demand equation system that contains exclusively the three beverages and the total expenditure on alcohol and an unconditional demand equation system that includes total expenditures on all goods and the price index of all other goods.

The almost ideal demand system (AIDS) was developed by Deaton and Muellbauer (1980*a*), who assumed that

$$w_{it} = \alpha_i + \sum_{j=1}^{n} \beta_{ij} log(p_{jt}) + \gamma_i log(m_t/P_t)$$

where P_t = price index in period t defined by

$$log(P_t) = \alpha_0 + \sum_{i=1}^{n} \alpha_i log(p_{it}) + \tfrac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \beta_{ij} log(p_{it}) log(p_{jt}).$$

For this system, the homogeneity and symmetry restrictions are also given by $\Sigma_{j=1}^{n} \beta_{ij} = 0$ and $\beta_{ij} = \beta_{ji}$ $(i \neq j)$.

For both the Rotterdam system and the AIDS, the homogeneity and symmetry restrictions can be tested or imposed. Clements and Johnson (1983) tested the restrictions and could not reject them. Thom (1984) rejected the homogeneity and symmetry restrictions in his static AIDS model. Selvanathan (1988) and Jones (1989) did not test the restrictions.[20] Their estimates of the price elasticities are reproduced in table 4. One can see that the findings of Clements and Johnson, Selvanathan, and Jones are very similar. They found that beer, spirits, and wine are all demand inelastic, with beer the most inelastic and spirits the most elastic. On the other hand, the magnitude of Thom's estimates are more in line with Ornstein and Levy's benchmark. All these results suggest that the price-elasticity estimates obtained from the system approach are consistent with those derived from the single-equation approach, therefore adding to the robustness of the findings in the literature.

Of course, not all foreign work adopts the system approach to analyze alcohol

[19] *In addition to the homogeneity and symmetry restrictions, demand theory also implies several adding-up restrictions. However, these adding-up restrictions are normally automatically satisfied by design, so they do not impose any real restriction on the estimation of the model (see Deaton and Muellbauer 1980b).*

[20] *Deaton and Muellbauer's (1980b) review of the literature shows that the homogeneity and symmetry restrictions are consistently rejected by data. They provided an excellent assessment of the strengths and weaknesses of the Rotterdam and AIDS models and other system approaches to demand analysis.*

TABLE 4

Estimates of the price elasticities of demand of alcoholic beverages: Aggregate data

Authors	Place, time period, type of data	Model and estimation method	Price elasticity			
			Beer	Liquor	Wine	Alcohol
Cook (1981)	U.S., 1960–75, time-series cross-sectional (23 States)	Quasi-experimental method		–1.6*		
Cook and Tauchen (1982)	U.S., 1962–77, time-series cross-sectional (30 States)	Two-step generalized least squares		–1.8†		
Clements and Johnson (1983)	Australia, 1955–77, time-series	Rotterdam system	–0.36† (uncon-ditional)	–0.74† (uncon-ditional)	–0.43† (uncon-ditional)	
Duffy (1983)	U.K., 1963–78, time-series	Ordinary least squares and two-stage least squares	0.23* (OLS) 0.22 (2SLS)	–1.18† (OLS) –0.13 (2SLS)	–0.64* (OLS) –0.71* (2SLS)	
Schweitzer et al. (1983)	U.S., 1975, cross-sectional (35 States)	Ordinary least squares	–0.27 (reduced form) –0.47* (structural model)	–0.13 (reduced form) 0.35 (structural model)		
Tsolakis et al. (1983)	Australia, 1955–79, time-series	Ordinary least squares, maximum likelihood, and Houthakker-Taylor (HT) model	–0.22† (log-linear model) –0.23 (short-run HT model) –0.20 (long-run HT model)		–0.27* (log-linear model) –0.43* (short-run HT model) –1.35* (long-run HT model)	
Levy and Sheflin (1983, 1985)	U.S., 1940–80, time-series	Ordinary least squares				–0.5†
Thom (1984)	Ireland, 1969–80, time-series	Almost ideal demand system	–0.68‡	–1.42‡	–1.6‡	
Ornstein and Hanssens (1985)	U.S., 1974–78, time-series cross-sectional (50 States and D.C. for spirits, 48 States and D.C. for beer)	Ordinary least squares	–0.12* (all States) –0.17* (license States)	–0.92† (all States) –0.66† (license States)		

TABLE 4 *(continued)*

Estimates of the price elasticities of demand of alcoholic beverages: Aggregate data

Authors	Place, time period, type of data	Model and estimation method	Price elasticity Beer	Liquor	Wine	Alcohol
Uri (1986)	U.S., 1982, cross-sectional (50 States and D.C.)	Maximum likelihood	–1.07†	–1.21†	–0.88†	
Adrian and Ferguson (1987)	Canada, 1958–81, time-series cross-sectional (7 provinces)	Ordinary least squares	–0.37† (domestic) –0.84† (imported)	–0.05 (domestic) –0.96† (imported)	–0.61† (domestic) –1.27† (imported)	
Godfrey (1988)	U.K., 1956–80, time-series	Ordinary least squares	–1.05 (linear Duffy model)	–0.56† (log-linear McGuinness model)	–2.67† (log-linear Duffy model)	
Selvanathan (1988)	U.K., 1955–85, time-series	Rotterdam system	–0.20* (uncon-ditional model)	–0.79† (uncon-ditional model)	–0.49† (uncon-ditional model)	
Jones (1989)	U.K., 1964–83, time-series	Almost ideal demand system	–0.27† (habit model) –0.40† (non-habit model)	–0.95† (habit model) –0.79† (non-habit model)	–0.77† (habit model) –0.94† (non-habit model)	
Nelson (1990)	U.S., 1982, cross-sectional (48 States and D.C.)	Ordinary least squares	–0.56†	–0.89†	–1.86†	

Notes: Some studies do not report the significance levels of the elasticity estimates. In these cases, the significance levels reported here referred to those of the corresponding price coefficients.

* *Significant at 10-percent level.*

† *Significant at 5-percent level.*

‡ *Significance level not reported and cannot be inferred.*

demand. There are several studies, each with a different focus, that employed the conventional single-equation approach. Duffy (1983) investigated the demand for beer, spirits, and wine in the United Kingdom with a special emphasis on the effect of advertising expenditures. Tsolakis et al. (1983) focused on the

dynamic aspects of alcohol demand in Australia. Adrian and Ferguson (1987) investigated the differences between the price sensitivities of Canadian and imported alcoholic beverages. Godfrey (1988) analyzed the effect of licensing on the consumption of alcohol in the United Kingdom. Table 4 contains some of the estimates of the price elasticities from these four studies. In general, the price of beer is found to be the most inelastic. There does not appear to be a consistent relationship between the price elasticities of spirits and wine. The most disturbing results are those of Godfrey, who found that the estimates are highly sensitive to the models and specifications. The price elasticities vary over a wide range of values, from –0.56 to –3.9 for spirits and from –0.09 to –2.05 for wine. This is a fundamental problem that will be addressed in the section entitled General Discussion.

STUDIES BASED ON INDIVIDUAL DATA

Kendell, de Roumanie, and Ritson (1983)

The work by Kendell et al. (1983) appears to be the first study on the effect of price on consumption at an individual level. The main purpose was to find out whether heavy drinkers reduce their consumption when the price of alcoholic beverages increases. Their analysis was based on a natural experiment performed in 1981 in Scotland. In 1978–79, as part of a World Health Organization survey of drinking habits and attitudes concerning alcohol, a sample of 608 men and 399 women living in the Lothian region (Edinburgh and its hinterland) aged above 17.5 was asked what they had drunk in the previous 7 days.

In March 1981, for the first time in more than 30 years, the price of alcoholic beverages increased faster than the retail price index, due to a sharp rise in the excise tax on alcoholic beverages and increases in prices imposed by the brewers themselves. A survey was immediately carried out in the fall of 1981 to reinterview respondents who had claimed to have at least one alcoholic drink in the 7-day period in the original survey. This natural experiment provided an unprecedented opportunity to perform a before/after comparison to find out the effect of price increase on the consumption of alcoholic beverages. Between the two surveys (designated T1 and T2), the price of alcoholic beverages had risen 61 percent and the retail price index 52 percent, so the real price of alcoholic beverages had risen 9 percent. One of the main results is described in table 5.[21]

Compared with the range of estimates in the literature, the absolute values of the price elasticities were exceptionally large, indicating that demand was highly elastic. The impact of the price increase was strongest among the heaviest drinkers (more than 50 units/week for men and

[21] *Dividing the D score in table 3 in Kendell et al. (1983) by the mean consumption at T1, one obtains the percentage change in mean consumption between T1 and T2. The price elasticity is then obtained by dividing this number by 9 percent (the percentage change in price between T1 and T2).*

more than 10 units/week for women). Surprisingly, the lightest drinkers increased their consumption. Kendell et al. (1983, p. 369) considered these results misleading because of the regression to the mean effect. They argued the problem is that

> . . . most people's drinking varies from week to week. Some of the men drinking only 1 to 10 units/week at T1 [the first survey] would normally have drunk more than this and some of those drinking over 51 units would normally have drunk less. So when, at T2 [the second survey], they are interviewed at the end of a more typical week light drinkers appear to be drinking more and heavy drinkers to be drinking less.

Another problem is that for financial reasons, the second survey did not reinterview respondents who happened not to have had a single drink in the first survey. By adjusting for these two problems, they concluded that "the reduction was distributed across the whole range of consumption with a (statistically insignificant) tendency for the heaviest male drinkers to cut down rather more than other people" (Kendell et al. 1983, p. 376). Unfortunately, they did not detail their adjustment procedure. Furthermore, without access to the data, it is impossible to evaluate their results. Nevertheless, they provided some other findings to strengthen their conclusion that heaviest drinkers were more likely than other people to reduce their consumption. For instance, the heaviest drinkers were more likely than others to respond that they drank less in the second survey than in the first one. Heavy drinkers also experienced the largest reduction in the number of adverse effects related to alcohol consumption.

TABLE 5

Kendell, de Roumanie, and Ritson estimates of price elasticities for each drinking group

Consumption (drinks per week)	Sample size	Price elasticity
Men		
1–10	137	7.87
11–30	117	0.65
31–50	41	–3.07
Over 50	34	–5.85
Total	329	–1.95
Women		
1–5	74	2.99
6–10	31	–2.39
Over 10	29	–4.08
Total	134	–2.19

There are several shortcomings in the Kendell et al. analysis. First, the 7-day period offers a time interval too short for people to describe their real drinking behaviors. The consumption variable is likely to be contaminated with serious measurement errors. Furthermore, the fact that respondents with zero consumption in the first survey were not reinterviewed poses a serious selection bias in the sample. Therefore, the design of the two surveys undermines the authors' conclusions. Second, although the data were collected at the individual level, most of

the calculations and estimates reported are aggregate numbers or percentages. In addition, they only studied the relationship between two variables (e.g., consumption and income, consumption and social class, consumption and employment status, consumption and marital status) and never put them together. A more appropriate approach would be to control the variables simultaneously, as in a standard multiple regression model.

Grossman, Coate, and Arluck (1987) and Coate and Grossman (1988)

By far the most extensive studies on the demand for alcoholic beverages using U.S. individual data are Grossman et al. (1987) and Coate and Grossman (1988). Their main objectives were to study the price sensitivity of youth alcohol demand and the effectiveness of minimum drinking age laws. The two studies had the same objectives and used the same estimation method, and the analyses were similar in many respects. Both studies focused only on youths between the ages of 16 and 21.[22] The only main difference is the source of data. Grossman et al. obtained the price data from an experimental pricing program conducted by the Bureau of Labor Statistics, while Coate and Grossman obtained the price of beer from Ornstein and Hanssens (1985) and the price of liquor (Seagram's 7-Crown) from the Distilled Spirits Council of the United States. For alcohol consumption data, Grossman et al. used the first National Health and Nutrition Examination Survey (NHANES I), whereas Coate and Grossman used the second one (NHANES II). NHANES I and NHANES II cover the periods 1971–74 and 1976–80, respectively, and they also differ in the survey questions on drinking behavior.

Grossman et al. created three measures of drinking from NHANES I: participation, frequency, and quantity (number of drinks on a typical drinking day). The explanatory variables are prices, minimum legal age, and individual background variables such as sex and family income. Participation is treated as a dichotomous variable, and three separate dichotomous logit models are estimated, one for each type of beverage participation. Both the frequency of drinking (which consists of five outcomes) and the quantity of drinking (which is converted into four categories from the original continuous variables) are estimated by multinomial logit models. They focused on beer and liquor, because the results for wine were not satisfactory. All the cross-price effects were found to be insignificant.

The price elasticities obtained from the logit models have a different meaning than those obtained from OLS regressions. Consider, for instance, the multinomial logit model for the quantity of drinking (number of drinks on a typical drinking day), which is divided into three outcome categories: six or more, three to five, and one to two. A rise in the price of beer will change the probability of each outcome category, but it does not tell how

[22] *The Grossman et al. analysis is much more elaborate, whereas Coate and Grossman only discuss the results for beer.*

the number of drinks will change. The price elasticity measures the percentage change in probability, not the percentage change in quantity, resulting from a 1-percent change in price. The inability to produce estimates of the price elasticity of demand is a result of the categorization of the drinking variables. Accordingly, their estimates of price elasticities cannot be compared directly with others in the literature.

For the participation logit models, the price effects are negative and statistically significant for beer and liquor. For the frequency and the quantity multinomial logit models, the price effects of all the outcome categories are negative for beer (but a majority of the estimates are statistically insignificant) and the price effects of most of the outcome categories for liquor are negative and significant. These results provide weak evidence that heavy drinkers (especially for beer) are more sensitive to price changes than are moderate drinkers. Using tables from the Grossman et al. (1987) article, one can obtain a rough estimate of the price elasticities of each outcome category.[23]

Table 6 shows that most of the price elasticities exceed unity, which suggests that the demand for beer and liquor are sensitive to price changes, even for heavy drinkers. Two exceptions are the positive elasticities for heavy liquor drinkers (+3.38 and +3.83). These two estimates are counterintuitive and are likely to be caused by the small samples of youth in these categories (1.4 and 1.1 percent, respectively), which may render the estimates unreliable. As discussed previously, these are not the conventional elasticities, and they do not indicate how the quantities of alcohol consumed will vary with changes in prices.

TABLE 6

Grossman, Coate, and Arluck estimates of price elasticities of participation, frequency, and quantity (youths ages 16–21)

	Beer	Liquor
Participation	–1.54	–4.07
Frequency		
Every day	–3.29	none
2–3 times/week	–0.51	3.38
1–4 times/month	–1.90	–5.26
4–11 times/year	–1.45	–4.31
Never	0.73	0.60
Quantity		
6 or more	–3.05	3.83
3–5	–1.58	–7.66
1–2	–1.19	–3.78
None	0.72	0.61

One way to think about the combination of quantity and frequency elasticities comes from the usual model in which expected demand is divided into two components, the probability of participation and the conditional quantity of participation. Thus, $E(Q) = \pi Q$, where π is the

[23] *The elasticities are calculated as follows: Tables 10–12 in Grossman et al. (1987) are calculated by assuming a $0.10 increase in the price of beer and a $0.30 increase in the price of liquor. Using the mean prices of beer and liquor in table 2, the percentage increase in price becomes 7 percent (0.1/1.426) for beer and 5.7 percent (0.3/5.275) for liquor. The elasticities are then obtained by dividing the percentage changes in Tables 10–12 in Grossman et al. (1987) by 7 for beer and 5.7 for liquor.*

probability of participation and Q is the conditional quantity consumed. It is easy to show that the elasticity of $E(Q)$ is the sum of the elasticities of π and Q with respect to price. One could combine the responsiveness estimates in Coate and Grossman (1988) by computing the expected amount of consumption for a randomly drawn individual at one price, perturbing that price slightly, and recomputing the expected consumption. The "aggregate" demand elasticity would represent the change in expected consumption relative to the change in price. In this way, one could use quantity-frequency estimates such as those of Coate and Grossman (1988) or Grossman et al. (1987) to derive an elasticity estimate comparable in style, if not in underlying estimation approach, to other estimates in the literature.

Coate and Grossman (1988) only report the estimates of a multinomial logit model for beer consumption frequency.

TABLE 7

Coate and Grossman estimates of price elasticities of frequency of consumption (youths ages 16–21)

Frequency	Model excluding religion variables	Model including religion variables
4–7 times/week	–1.18	–0.53
1–3 times/week	–0.59	–0.48
Less than once/week	0.05	–0.20
None	0.65	0.53

The price effects are negative and generally significant.[24] However, when religion variables are included in the model, all the price effects become insignificant, although they remain negative. The price elasticities are shown in table 7.

Although the definitions of the frequency categories in tables 6 and 7 differ, it is clear that the absolute values of the price elasticities are much smaller than those of Grossman et al. This question is not addressed in the Coate and Grossman paper.

Heien and Pompelli (1989)

Heien and Pompelli's study was motivated by the concern that most studies on the demand for alcoholic beverages in the United States are dated, and the available studies do not treat all beverages (alcoholic and nonalcoholic) simultaneously and do not examine demographic effects. They attempted to remedy these shortcomings by applying the AIDS (Deaton and Muellbauer 1980*a*) to estimate a complete demand system for beer, distilled spirits, wine, and selected nonalcoholic beverages.[25] The data set was obtained from the 1977–78 Household Food Consumption Survey collected by the U.S. Department of Agriculture, which contains data on prices and quantities for over 1,000 food items consumed by 14,000 households. Their findings are given in table 8.

Heien and Pompelli found that the three alcoholic beverages are weak complements to each other. This is exactly oppo-

[24] *The price effect for the category "less than once a week" is insignificant.*

[25] *The nonalcoholic beverages included in their study are coffee, tea, soft drinks, juice, and milk.*

site to what Uri (1986) found. The demand for beer, spirits, and beverages were all inelastic, with spirits the most inelastic and beer the closest to unitary elastic.[26] The ranking of the price elasticities is exactly the opposite of the Ornstein and Levy benchmark. They also found that all the income elasticities, including those of nonalcoholic beverages, are negative. This is an unusual result because it is difficult to conceive that all beverages, alcoholic and nonalcoholic, are inferior goods.

One of the major problems with their estimation procedure is that censoring is ignored. With individual data, there are households with zero expenditures for some beverages. It is well-known that ignoring the problem of censoring will give biased and inconsistent estimates, even in a simple OLS regression (Maddala 1983). Without controlling for censoring, it is unlikely that the standard statistical properties of the AIDS estimates will be preserved. Hence, it is not clear how much confidence one should have in the estimates.[27]

Atkinson, Gomulka, and Stern (1990)

In the United Kingdom, household expenditure on alcoholic beverages constitutes a substantial portion of total household spending (about 7.5 percent). The aim of Atkinson et al. (1990) was to model how observed characteristics affect household expenditure on alcoholic beverages and the incidence of zero expenditure. They used the U.K. Family Expenditure Surveys from the period April 1970 to December 1983. There are 68,854 households in the sample, and about 20 percent of them have zero recorded expenditure for alcohol. To control for these censored data, they estimated a Tobit model with a more general error term:

TABLE 8

Heien and Pompelli estimates of price elasticities

	Beer	Spirits	Wine
Beer	–0.84	–0.02	–0.06
Spirits	–0.09	–0.50	–0.23
Wine	–0.21	–0.31	–0.55

Note: Each number in the table represents the elasticity of the column variable with respect to the price of the row variable.

$$w = max\ \{0, \alpha + \beta log(m/\pi) + \delta[log(m/\pi)]^2 + \gamma log(p/\pi) + \varepsilon\} \qquad (3)$$

where w = share of expenditure on alcohol, m = total household expenditure, p = price of alcohol, and π = retail price index. Since the distribution of observed expenditure shares is positively skewed (skewed to the right), they argued that the normality assumption of the error term ε (a standard assumption in the econometrics literature) may not be appropriate. They

[26] *They did not report whether the estimates are statistically significant.*

[27] *This problem was also emphasized in Deaton (1986). He maintained that "the problem of dealing appropriately with zero expenditures is currently one of the most pressing in applied demand analysis. We do not have a theoretically satisfactory and empirically implementable method for modelling zeroes for more than a few commodities at once. Yet all household surveys show large fractions of households reporting zero purchases for some goods" (Deaton 1986, p. 1809).*

therefore assumed a gamma distribution (Γ) for ε, which includes the normal distribution as a special case. They found that the gamma-Tobit fits the data much better than the Tobit with normal errors.

With the Tobit model, there are at least two ways to calculate the price elasticities of demand. The expenditure function in equation 3 implies that the price elasticity of demand is $1 - \gamma/w$ for households with positive alcohol expenditures. If one considers both zero and positive expenditures (the entire sample), the price elasticity is $1 - \gamma[1 - Pr(w > 0)]/E(w)$. This is because the expected share w for the whole sample is given by

$$E(w) = [1 - Pr(w > 0)]\{\alpha + \beta log(m/\pi) + \delta[log(m/\pi)]^2 + \gamma log(p/\pi)\} \quad (4)$$

where $Pr(w > 0) = 1 - \Gamma\{-\alpha - \beta log(m/\pi) - \delta[log(m/\pi)]^2 - \gamma log(p/\pi)\}$ = the probability that the expenditure share on alcohol is positive and $\Gamma(.)$ is the cumulative distribution function of the gamma distribution. For convenience, let us call the first method (the one that uses $1 - \gamma/w$) the conditional model and the second method [the one that uses $1 - \gamma[1 - Pr(w > 0)] \div E(w)$] the unconditional model.[28] For the conditional model, over 70 percent of drinkers have a price elasticity of –1.3 to –1.0, and there are some 10 percent with a price elasticity below –2.0. For the unconditional model, over 99 percent of drinkers have price elasticities between –1.15 and –1.11.[29] Unlike the conditional model, the distribution of the price elasticities for the unconditional model is very concentrated around the mean. The means of the price elasticities calculated from tables A.3 (a) and (b) in the Atkinson et al. paper are –1.42 for the conditional model and –1.12 for the unconditional model. Thus, they found evidence of an elastic demand for alcohol, which differs from most other studies. One possible explanation is that the survey omitted a minority of heavy drinkers who may have had a more inelastic demand for alcohol than those sampled in the survey.

Kenkel (1991)

Although the main focus in Kenkel (1991) was drunk driving, he also estimated a demand model for heavy drinking. The data set comes from the Health Promotion and Disease Prevention Supplement to the 1985 Health Interview Survey. There are 12,385 males and 16,686 females in the sample. The dependent variable is HEAVY, which is the number of days in the past year the individual reports having had five or more drinks of any alcoholic beverage. This variable measures the frequency of heavy drinking. The price variable is a State-level weighted average of the average beer, wine, and liquor prices provided by Nelson (1990). There are two groups of explanatory variables: State deterrence laws on drunk driving (e.g., breath test, sobriety checkpoints) and alcohol control

[28] *Atkinson et al. (1990) interpreted the conditional model as a "fixed effect" model and the unconditional one as a "random transitory component" model.*

[29] *The distribution of price elasticities is induced by the distribution of expenditure shares.*

policy variables (e.g., monopoly versus license States, minimum drinking age). The model is a standard Tobit:

$$HEAVY_i = max\ \{0, X_i\beta + u_i\} \qquad (5)$$

where X_i is a vector of explanatory variables, β is a vector of parameters, and u_i is an independent and identically distributed error term with a normal distribution. The price elasticities of demand obtained from the unconditional model and evaluated at the means of the variables are –0.48 for males and –1.07 for females.

Kenkel also estimated the model for a subsample of young adults (age 21 and below) and found that the price elasticities are –0.58 for males and –2.89 for females, which suggests that young adults have a more elastic demand for alcohol than older people. He regarded the results as evidence that young adults' demands for alcohol are more sensitive to price changes than those of older age groups. There are two problems with this interpretation. First, the estimate of the price coefficient for young males is statistically insignificant, hence the price elasticity derived from it should not be taken with much confidence. Second, the price elasticities obtained from the whole sample are not comparable with those obtained from the subsample because the models are different and the variables do not have the same values.[30]

GENERAL DISCUSSION

Compared with the pessimism appearing in Ornstein (1980), Ornstein and Levy (1983, p. 344) ended their survey with a more positive recommendation for future research on the demand for alcoholic beverages:

> There is clearly room for further work on demand estimation of beer, wine, and distilled spirits. First, there is room for improvement in econometric technique. Second, the quality of the data used is poor in many cases, especially with regard to price indices. Third, of the three beverages, far less is known about the demand for wine. There is a need for additional studies on wine consumption. Fourth, none of the United States' studies have examined the parameters of demand within incidence of drinking groups. There is a need to develop data by drinking groups, for the real issue is what effect changes in the determinants of demand will have on occasional, light, moderate, and heavy drinkers.

[30] *Recall that for a logarithmic demand equation, the price elasticity is directly obtained from the OLS estimate of the price coefficient. For a linear demand equation, the price elasticity depends on the price coefficient and the values of the price and quantity variables (usually the mean values of these variables are used). However, for a Tobit model, the price elasticity depends on the price coefficient and all the other coefficients and variables in the model. Hence, unless one controls for the differences in all the variables, one cannot compare the price elasticities between two Tobit models because the differences between the price elasticities may just come from variations in the values of the variables.*

Researchers in the past decade have begun to carry out some of these recommendations. On the whole, our review shows that the findings of the studies that employ aggregate data are in accord with those reviewed in Ornstein and Ornstein and Levy and provide some support for the benchmark price elasticities of –0.3, –1.0, and –1.5 for beer, wine, and distilled spirits, respectively. The most significant advance, however, is in the use of individual data in alcohol studies that attempt to gauge the price effects on various groups of drinkers. As summarized in table 9, studies employing individual data appear to find more elastic demand for alcoholic beverages. Whether the discrepancies are due to differences in the nature of the data (aggregate versus individual) or in the estimation methods remains to be addressed.

There has not been much progress toward the second and third points listed in Ornstein and Levy's recommendations. The quality of price data, especially for wine, has not improved much in the studies reviewed here. Almost all domestic studies ignore the demand for wine. Researchers either could not obtain the price data for wine (Ornstein and Hanssens 1985) or could not obtain satisfactory estimation results for wine (Grossman et al. 1987). Some researchers tried to avoid the problem by just examining the demand for alcoholic beverages as a whole and by calculating price indices from expenditure and quantity data (Levy and Sheflin 1983, 1985). However, there are some problems with measuring price by the value of expenditure per unit of alcohol. For instance, Johnson et al. (1990) showed that it is possible for the price index to fall even when there is an increase in all the prices of beer, distilled spirits, and wine.

The estimates for the cross-price elasticities continue to puzzle researchers. Some studies (e.g., Uri 1986) found that beer, spirits, and wine are all substitutes, while others (e.g., Heien and Pompelli 1989) found that they are all complements, and yet others (e.g., Selvanathan 1988) found that they are all independent commodities. It is not clear whether these contradictory results are caused by the quality of the price data or by some methodological problems. It should be emphasized that alcohol demand studies are not the only demand studies that cannot produce robust estimates of cross-price elasticities. In fact, robust estimation of cross-price elasticities is generally a difficult task shared by demand studies of all the other commodities (see Deaton and Muellbauer 1980*b*, p. 79).

Ornstein and Levy's first recommendation has been implemented in some of the studies. Examples are the applications of the Granger-Sims causality test in Cook and Tauchen (1982) and the gamma-Tobit in Atkinson et al. (1990). However, there are some fundamental econometric problems that should have been addressed in the literature. We provide one example here.

A major problem in the empirical work employing aggregate data is the choice of functional forms for the OLS regression equation (the demand function). The two main competing specifications are the linear and the logarithmic

TABLE 9

Estimates of the price elasticities of demand of alcoholic beverages: Individual data

Authors	Place, time period, type of data	Estimation method	Price elasticity Beer	Liquor	Wine	Alcohol
Kendell et al. (1983)	U.K., 1978–82, individual	Before/after comparisons of sample means				–1.95 (male) –2.19 (female)
Grossman et al. (1987)	U.S., 1971–74, individual (ages 16–21)	Multinomial logit	–1.54† (participation) –3.29 (frequency) –3.05* (quantity)	–4.07† (participation) none (frequency) 3.83 (quantity)		
Coate and Grossman (1988)	U.S., 1976–80, individual (ages 16–21)	Multinomial logit	frequency: –0.53 (with religion variables) –1.18† (without religion variables)			
Heien and Pompelli (1989)	U.S., 1977–78, individual	Almost ideal demand system (AIDS)	–0.84	–0.5†	–0.55†	
Atkinson et al. (1990)	U.K., 1970–83, individual	Gamma–Tobit				–1.42† (conditional model) –1.12† (unconditional model)
Kenkel (1991)	U.S., 1985, individual	Tobit				–0.48† (male) –0.58 (young male) –1.07† (female) –2.89† (young female)

Notes: Some studies do not report the significance levels of the elasticity estimates. In these cases, the significance levels reported here referred to those of the corresponding price coefficients.

* *Significant at 10-percent level.*

† *Significant at 5-percent level.*

demand functions. The logarithmic demand function is especially popular in demand studies because it generates constant elasticities that are convenient to interpret. Since there is no a priori theory indicating which functional form is a more appropriate choice, most studies report estimates for both linear and logarithmic demand functions (Prest 1949; Lau 1975; Johnson and Oksanen 1974, 1977; Walsh and Walsh 1970). Yet empirical results indicate that the estimates can be very sensitive to different specifications of the demand equation. For example, Godfrey's (1988) estimates show that depending on the functional form used, the price elasticity of distilled spirits could range from –0.56 to –3.9. At the time when those earlier alcohol studies were carried out, no econometric tests were available to test for functional forms. However, recent advances in econometrics have provided a variety of tests that can be used to shed light on the problem. In particular, the tests proposed by Davidson and MacKinnon (1981, 1984; see also MacKinnon et al. 1983) are conceptually the simplest and computationally the easiest to use.

Godfrey (1988) applied the Davidson and MacKinnon (1981) J-test in her study of the effect of licensing on the demand for alcohol in the United Kingdom. The problem with the J-test is that it is designed to test models with the *same dependent variable.* In other words, the J-test is not applicable when the models do not contain the same dependent variable, for example, when the dependent variable in one model is in logarithmic form and the other one is not. This is a serious deficiency because the two competing models in the alcohol literature are usually the linear model and the logarithmic model. Godfrey's analysis clearly exemplifies the problems and limitations of the J-test. In these cases, one can use the test developed by MacKinnon et al. (1983).

Consider the following linear and semilogarithmic demand functions:

$$H_0: q_i = \sum_j a_j X_{ij} + \sum_j b_j Z_{ij} + u_i \qquad (6)$$

$$H_1: log(q_i) = \sum_j c_j log(X_{ij}) + \sum_j d_j Z_{ij} + v_i \qquad (7)$$

where q_i = consumption of alcoholic beverages, X_{ij} = explanatory variables with which logarithms can be taken (e.g., price, income), Z_{ij} = explanatory variables with which logarithms cannot be taken (e.g., dummy variables, variables with negative values). The MacKinnon-White-Davidson test can be used to test H_0 against H_1, or H_1 against H_0. The steps are as follows:

(i) Obtain the OLS estimates $(\hat{a}_j, \hat{b}_j, \hat{c}_j, \hat{d}_j)$ for equations 6 and 7.

(ii) To test H_0 against H_1, estimate θ_j, ϕ_j, and α of the following ordinary linear regression:

$$q_i = \sum_j \theta_j X_{ij} + \sum_j \phi_j Z_{ij} + \alpha \left[\sum_j \hat{c}_j log(X_{ij}) + \sum_j \hat{d}_j Z_{ij} - log(\sum_j \hat{a}_j X_{ij} + \sum_j \hat{b}_j Z_{ij})\right] + w_i$$

Perform a *t*-test on $\hat{\alpha}$. If $\hat{\alpha}$ is not significantly different from 0, then the semi-

logarithmic model is rejected in favor of the linear model.

(iii) To test H_1 against H_0, estimate θ_j, ϕ_j, and α of the following ordinary linear regression:

$$\log(q_i) = \sum_j \theta_j \log(X_{ij}) + \sum_j \phi_j Z_{ij} + \alpha[(\sum_j \hat{a}_j X_{ij} + \sum_j \hat{b}_j Z_{ij}) - \exp(\sum_j \hat{c}_j \log(X_{ij}) + \sum_j \hat{d}_j Z_{ij})] + w_i$$

Perform a t-test on $\hat{\alpha}$. If $\hat{\alpha}$ is not significantly different from 0, then the linear model is rejected in favor of the semilogarithmic model.

This test for functional forms is computationally extremely easy to use: only three OLS regressions and two t-tests are needed. A potential problem with the test is that it is possible to find that both the linear and the logarithmic models are rejected (or neither one is rejected) after carrying out the tests in steps ii and iii.

Alternative approaches have relied on a generalized Box-Cox transform of the dependent variable, specifying a general model where $z = y^{\lambda}/\lambda$, which includes the linear form when $\lambda = 1$ and the logarithmic form asymptotically as $\lambda = 0$. It also contains intermediate power transformations, many of which are well-approximated (for example) by the Γ distribution. A general maximum likelihood search over λ determines the best model. This approach is more complicated than the previous one because it requires nonlinear estimation, although a grid search over λ may yield good results with linear models conditional on the value of λ. Thus far the only application (to our knowledge) of this approach in the alcohol literature is the one by Tsolakis et al. (1983). They found that the method works well in the wine equations, but it fails in the beer equations because they could not locate a global maximum on the likelihood surface. As a result, they had to abandon the approach and resort to those commonly analyzed functional forms for the demand equations for beer.

For studies based on individual data, there is even more room for improvement in econometric technique because the data structures are usually more complicated. We give two examples here.[31] Grossman et al. (1987) converted the number of drinks consumed on a typical day from a continuous to a categorical variable to allow for the possibility of nonlinearity and nonmonotonicity. They found that the estimates for different groups of drinkers are different and inferred that there are nonlinear effects, but they did not test whether the differences are statistically significant. To substantiate their claims, one could use the Wald test, the likelihood ratio test, or the Lagrange multiplier test to examine whether the estimates across different groups of drinkers (heavy, moderate, and light) are statistically different.

A second example is the treatment of the relationship among the three drinking variables that are frequently found in surveys of drinking behavior: participation, frequency, and quantity (number of drinks on a typical drinking day).

[31] *These two examples are addressed in more detail in our current work in progress on the price sensitivity of alcoholic beverages using the 1983 National Health Interview Survey.*

Grossman et al. (1987) assumed that they are independent random variables and treated them separately by estimating a logit (or multinomial logit) model for each variable. On the other hand, Kenkel (1991) assumed that drinking participation and drinking quantity are related and used the same equation to model the two variables (Type I Tobit).[32] Clearly, future research should test (instead of assume) the relationships among the variables.

The emergence of studies employing individual data raises many econometric problems that need to be addressed in future research. In addition to the two examples cited, the choice between logit and Tobit models, the restrictive assumption underlying the multinomial logit model,[33] the distributional assumption on the error term of the Tobit model, and the choice between conditional and unconditional models in calculating elasticities are just a few of the problems that require further work. Recent developments in hypothesis testing (Pagan and Vella 1989) and semiparametric estimations (Newey 1990) should be able to contribute to the solution of these problems in alcohol demand studies.

CONCLUSION

We note in conclusion that even as the literature has moved to much more disaggregated data, the estimates in the literature contain a wide degree of disagreement about the price responsiveness of demand for alcoholic beverages. The problems remaining in the literature will be partly addressed by improvements in the econometric estimation techniques. Ultimately, the underlying quality of the data will determine the value of the studies.

As we learn more about the econometric properties of various demand estimators, we will gain better insight into the underlying structure of demand and will also probe the limits of utility of existing data sources. At that point, it will be important to reassess the desirability of undertaking original data collection on consumer behavior that includes not only quantity and frequency of consumption (by type of beverage) but also direct measure of important covariates, including prices of beverages confronting individual consumers. Original data gathering also seems the only realistic way to understand clearly how behavioral responses might differ by type of drinker (e.g., moderate versus heavy), a concept that may prove important in public policy discussions of optimal taxation and other relevant issues.

ACKNOWLEDGMENTS

We are grateful to Philip Cook, Michael Grossman, John Mullahy, and especially Greg Bloss and Mike Hilton for their helpful comments and suggestions. This research was supported by National Institute on Alcohol Abuse and Alcoholism grant AA08393.

[32] *Atkinson et al. (1990) made essentially the same assumption, although their variable is expenditure share and not drinking quantity.*

[33] *The most serious drawback of the multinomial logit model is that it implies the independence of irrelevant alternatives property (Maddala 1983).*

REFERENCES

Adrian, M., and Ferguson, B.S. Demand for domestic and imported alcohol in Canada. *Appl Econ* 19:531–540, 1987.

Amemiya, T. *Advanced Econometrics.* Cambridge: Harvard University Press, 1985.

Atkinson, A.B.; Gomulka, J.; and Stern, N.H. Spending on alcohol: Evidence from the Family Expenditure Survey 1970–1983. *Econ J* 100:808–827, 1990.

Clements, K.W., and Johnson, L.W. The demand for beer, wine, and spirits: A systemwide analysis. *J Bus* 56:273–304, 1983.

Coate, D., and Grossman, M. Effects of alcoholic beverage prices and legal drinking ages on youth alcohol use. *J Law Econ* 31:145–171, 1988.

Cook, P.J. The effect of liquor taxes on drinking, cirrhosis, and auto accidents. In: Moore, M.H., and Gerstein, D., eds. *Alcohol and Public Policy: Beyond the Shadow of Prohibition.* Washington, DC: National Academy of Sciences, 1981. pp. 255–285.

Cook, P.J. Alcohol taxes as a public health measure. *Br J Addict* 11:245–250, 1982.

Cook, P.J., and Tauchen, G. The effect of liquor taxes on heavy drinking. *Bell J Econ* 13: 379–390, 1982.

Davidson, R., and MacKinnon, J.G. Several tests for model specification in the presence of alternative hypothesis. *Econometrica* 49:781–793, 1981.

Davidson, R., and MacKinnon, J.G. Model specification tests based on artificial linear regressions. *Int Econ Rev* 25:485–502, 1984.

Deaton, A., and Muellbauer, J. An almost ideal demand system. *Am Econ Rev* 70:312–326, 1980*a*.

Deaton, A., and Muellbauer, J. *Economics and Consumer Behavior.* New York: Cambridge University Press, 1980*b*.

Deaton, A. Demand analysis. In: Griliches, Z., and Intriligator, M.D., eds. *Handbook of Econometrics.* Vol. 3. New York: Elsevier, 1986. pp. 1767–1839.

Demsetz, H. Why regulate utilities? *J Law Econ* 11:55–65, 1968.

Duffy, M. The demand for alcoholic drink in the United Kingdom, 1963–78. *Appl Econ* 15:125–140, 1983.

Geary, P.T. The demand for petrol and tobacco in Ireland: A comment. *Econ Soc Rev* 4:201–207, 1973.

Godfrey, C. Licensing and the demand for alcohol. *Appl Econ* 20:1541–1558, 1988.

Greene, W.H. *Econometric Analysis.* New York: MacMillan, 1990.

Grossman, M.; Coate, D.; and Arluck, G. Price sensitivity of alcoholic beverages in the United States: Youth alcohol consumption. In: Holder, H.D., ed. *Control Issues in Alcohol Abuse Prevention: Strategies for States and Communities.* Greenwich, CT: JAI, 1987. pp. 169–198.

Heien, D., and Pompelli, G. The demand for alcoholic beverages: Economic and demographic effects. *S Econ J* 55:759–770, 1989.

Johnson, J.A., and Oksanen, E.H. Socio-economic determinants of the consumption of alcoholic beverages. *Appl Econ* 6:293–301, 1974.

Johnson, J.A., and Oksanen, E.H. Estimation of demand for alcoholic beverages in Canada from pooled time series and cross sections. *Rev Econ Stat* 59:113–118, 1977.

Johnson, J.A.; Oksanen, E.H.; Veall, M.R.; and Fretz, D. Alternative approaches to the measurement of consumption and price of alcoholic beverages, Canada, 1957–1983. *J Stud Alcohol* 51:82–85, 1990.

Jones, A.M. A systems approach to the demand for alcohol and tobacco. *Bull Econ Res* 41:85–105, 1989.

Kendell, R.E.; de Roumanie, M.; and Ritson, E.B. Effect of economic changes on Scottish drinking habits 1978–82. *Br J Addict* 78:365–379, 1983.

Kenkel, D.S. The Impact of State Laws on Drinking and Drunk Driving. Manuscript, Pennsylvania State University, 1991.

Keynes, J.M. *Essays in Biography.* New York: Macmillan, 1933.

Lau, H.H. Cost of alcoholic beverages as a determinant of alcohol consumption. In: Gibbins, R.J.; Israel, Y.; Kalant, H.; Popham, R.E.; Schmidt, W.; and Smart, R.G., eds. *Research Advances in Alcohol and Drug Problems.* Vol. 2. New York: Wiley, 1975. pp. 211–245.

Leamer, E., and Leonard, H. Reporting the fragility of regression estimates. *Rev Econ Stat* 65: 306–317, 1983.

Levy, D., and Sheflin, N. New evidence on controlling alcohol use through price. *J Stud Alcohol* 44:929–937, 1983.

Levy, D., and Sheflin, N. The demand for alcoholic beverages: An aggregate time-series analysis. *J Public Policy Marketing* 4:47–54, 1985.

MacKinnon, J.G.; White, H.; and Davidson, R. Tests for model specification in the presence of alternative hypotheses: Some further results. *J Econometr* 21:53–70, 1983.

Maddala, G.S. *Limited Dependent and Qualitative Variables in Econometrics.* Cambridge: Cambridge University Press, 1983.

Nelson, J. State monopolies and alcoholic beverage consumption. *J Reg Econ* 2:83–98, 1990.

Newey, W.K. Semiparametric efficiency bounds. *J Appl Econometr* 5:99–135, 1990.

Newman, P. Elasticity. In: Eatwell, J.; Milgate, M.; and Newman, P., eds. *The New Palgrave: A Dictionary of Economics.* New York: Macmillan, 1987.

Ornstein, S.I. Control of alcohol consumption through price increases. *J Stud Alcohol* 41:807–818, 1980.

Ornstein, S.I., and Hanssens, D. Alcohol control laws and the consumption of distilled spirits and beer. *J Consumer Res* 12:200–213, 1985.

Ornstein, S.I., and Levy, D. Price and income elasticities and the demand for alcoholic beverages. In: Galanter, M., ed. *Recent Developments in Alcoholism.* Vol. 1. New York: Plenum, 1983. pp. 303–345.

Pagan, A., and Vella, F. Diagnostic tests for models based on individual data: A survey. *J Appl Econometr* 4:S29–S59, 1989.

Phelps, C.E. Death and taxes: An opportunity for substitution. *J Health Econ* 7:1–24, 1988.

Pittman, D.J. Primary Prevention of Alcohol Abuse and Alcoholism: An Evaluation of the Control of Consumption Policy. St. Louis: Washington University Social Science Institute, 1980.

Pogue, T.F., and Sgontz, L.G. Taxing to control social costs: The case of alcohol. *Am Econ Rev* 79:235–243, 1989.

Prest, A.R. Some experiments in demand analysis. *Rev Econ Stat* 21:33–49, 1949.

Pudney, S. *Modelling Individual Choice: The Econometrics of Corners, Kinks, and Holes.* New York: Basil Blackwell, 1989.

Room, R. Alcohol control and public health. *Ann Rev Public Health* 5:293–317, 1984.

Schweitzer, S.O.; Intriligator, M.D.; and Salehi, H. Alcoholism: An econometric model of its causes, its effects and its control. In: Grant, M.; Williams, A.; and Plant, M., eds. *Economics and Alcohol.* London: Croom Helm, 1983. pp. 107–127.

Seeley, J.R. Death by liver cirrhosis and the price of beverage alcohol. *Can Med Assoc J* 83:1361–1366, 1960.

Selvanathan, E.A. Alcohol consumption in the U.K., 1955–85: A system-wide analysis. *Appl Econ* 20:1071–1086, 1988.

Simon, J.L. The price elasticity of liquor in the U.S. and a simple method of determination. *Econometrica* 34:193–205, 1966.

Stone, R. The analysis of market demand. *J Roy Stat Soc* 108:286–382, 1945.

Thom, D.R. The demand for alcohol in Ireland. *Econ Soc Rev* 15:325–336, 1984.

Tsolakis, D.; Riethmuller, P.; and Watts, G. The demand for wine and beer. *Rev Marketing Agric Econ* 51:131–153, 1983.

Uri, N.D. The demand for beverages and inter-beverage substitution in the United States. *Bull Econ Res* 38:77–85, 1986.

Walsh, B.M. The economics of alcohol taxation. In: Grant, M.; Williams, A.; and Plant, M., eds. *Economics and Alcohol.* London: Croom Helm, 1983. pp. 173–189.

Walsh, B.M., and Walsh, D. Economic aspects of alcohol consumption in the Republic of Ireland. *Econ Soc Rev* 12:115–138, 1970.

2

Taxation of Alcoholic Beverages

Philip J. Cook, Ph.D., and Michael J. Moore, Ph.D.

INTRODUCTION

The year 1991 marked the bicentennial of the first internal revenue measure adopted by Congress, an excise tax on whiskey (Hu 1950); a subsequent increase in that tax from 9 to 25 cents per gallon engendered an armed insurrection in the summer of 1794, the so-called Whiskey Rebellion (Godfrey 1990). Since then, increases in Federal alcohol excise taxes typically have been the *result* of a war rather than its cause,[1] but the issue remains contentious. As it was 200 years ago, the question now is not *whether* alcoholic beverages should be subject to higher taxes than other commodities, but just *how high* these taxes should be.

Interestingly, the substance of the debate over this issue has changed during the last decade. The principal impetus for raising tax rates has always been the quest for increased Government revenue, and that remains true today; however, in recent years there has been increasing attention to the public health benefits of increasing tax rates as well. For example, the 1987 *Economic Report of the President* (Council of Economic Advisors, pp. 187–188) noted that an increase in the Federal excise taxes on alcoholic beverages would reduce mortality rates from alcohol-related causes. The U.S. Department of Education's National Commission on Drug-Free Schools (1990) advocated an increase in alcohol excises as a deterrent to use by youths (p. 69), while the Public Health Service's recent (1990) report on national health promotion and disease prevention objectives, titled *Healthy People 2000*, advocated such an increase in order to reduce highway fatalities and cirrhosis deaths (p. 176). Outside of Government, the Center for Science in the Public Interest (CSPI) has publicized the case for excise taxes as a health promotion mechanism, and its National Alcohol Tax Coalition, representing 40 groups, has lobbied for a substantial increase on the same basis (Godfrey 1990).

The case for alcohol taxes as a health promotion measure may have helped per-

[1] *Federal excise taxes on liquor, beer, and wine were increased several times during each of the World Wars and again during the Korean War.*

suade Congress to increase beer and wine excises in 1990, the first such increase since 1951.[2] But the evidence that the tax level affects morbidity and mortality rates does not resolve the question of just how high taxes should be. Most commentators have asserted one of four standards to define the public interest: a historical/comparative standard, a fairness standard based on the external costs of drinking, a public health standard, and an economic efficiency standard.

It should be noted that while these principles are part of the political rhetoric of the alcohol tax debate, we make no claim concerning their actual importance in the political process. The political stakes are nontrivial; alcohol taxes generated $5.7 billion in Federal receipts in FY 1990 (under 1 percent of total Federal revenues) and a similar amount for State governments. Any proposal to raise tax rates engenders resistance from the alcohol industry and some consumers; for example, the Proposition 134 campaign in California to raise the State tax rates by a nickel per drink through a referendum measure on the 1990 ballot was defeated with the help of an industry-sponsored campaign. We will not attempt to analyze the political process by which tax rates are chosen, but rather will focus on applying alternative definitions of the public interest to the choice of tax rates.

We begin with a brief summary of the history of alcohol taxes and prices, concluding that current values are quite low in comparison with several decades ago. The next section analyzes the incidence of alcohol taxation, with particular focus on the implications of the "user fee" principle, that taxes should be set high enough to cover the social costs of drinking. Then we consider the case for alcohol taxes as a public health measure in greater detail and discuss the somewhat related notion that taxes should be used to correct for the negative externalities of drinking. In the next section, all of these approaches are brought to bear in an analysis of the proposal to adjust the structure of excise taxes across beverage types so as to equalize the tax per unit of ethanol. A final section offers some suggestions concerning future research directions.

THE HISTORICAL/COMPARATIVE STANDARD

Alcoholic beverages are subject to a variety of Federal, State, and local taxes. The most important of these are the special Federal and State excise taxes and the general State sales taxes. The excises are unit taxes, defined in terms of volume rather than value of the product. Currently, the Federal excise on spirits is $13.50 per proof gallon,[3] while beer is taxed at $18 per 31-gallon barrel and table wine at $1.07 per gallon. To facilitate comparison, we can calculate the average tax per ounce of ethanol from each beverage type: the result

[2] *In Title XI of the Omnibus Budget Reconciliation Act of 1990, Congress increased the tax on still wine (unfortified) from $.17 to $1.07 per gallon, the tax on beer from $9 to $18 per barrel, and the tax on distilled spirits from $12.50 to $13.50 per proof gallon. For beer and wine, these are the first changes since 1951. Congress had adopted a small increase in the distilled spirits tax in 1984.*

[3] *A "proof gallon" is equal to the volume of liquid that contains 64 ounces of pure alcohol.*

is 7 cents for wine, 10 cents for beer, and 21 cents for spirits (Eusterbrock 1990). In a subsequent section we discuss the arguments for bringing these rates into equality.

Maintaining the real value of the excise tax rates requires that the legislature increase the nominal value with inflation. However, Congress took no action to change alcohol excise taxes during the period 1951–90, except for a small increase in the liquor tax implemented in 1985. As a result of inflation during this period, by 1990, the Federal excise tax rates were only about 20 percent of their real value four decades earlier. The tax increases implemented in January 1991 reversed this long-term decline. As shown in figure 1, the real value of the Federal liquor tax was restored to its 1989 value and the beer tax to its 1978 value. The wine tax increase was more than large enough to restore its 1951 value.

The State excise taxes have also declined in real value over the last several decades. For example, while average State excise taxes on both beer and wine increased between 1966 and 1989 (to 22 cents per gallon for beer and 68 cents per gallon for wine), the real value of these taxes declined by over 50 percent during this period (CSPI 1990, p. 6).

The decline in the value of excise tax rates has contributed to a long-term decline in the prices of alcoholic beverages relative to other goods and services. Figures 2 and 3 show this price decline for the period since 1967. The average price of alcoholic beverages declined by 25 percent between 1967 and 1982 in comparison with the overall Consumer Price Index (CPI) and has remained roughly constant since then. This temporal pattern in prices represents an average over the three types of alcoholic beverages. When we look at these three price trajectories separately, as shown in figure 3, we see that spirits prices declined much more sharply than the prices of wine and beer during the 1960's and 1970's.

The historical standard for evaluating current excise tax rates can be expressed something like this: If in 1951, Congress saw fit to establish a tax rate equivalent in 1991 dollars to $8.40 per fifth of (80 proof) spirits, then why does our current Congress believe that the "right" tax is just $2.16 per fifth? Is there some reason why it is in the public interest to tax alcohol at a much lower rate now than then? Presuming the answer is no, then the case is made for restoring current taxes to the old level. There is no need in this formulation to agree on the optimal tax, but only on whether it should be going up or down over time.

Other sorts of comparisons may also be relevant in judging the appropriate tax rate. International and interstate comparisons reveal extraordinary variance in excise rates (Kay and Keen 1986, p. 86; CSPI 1990), reflecting such local factors as the strength of the temperance movement and the economic importance of wine or beer production. These differences have been seen as a barrier to the economic integration of the European Community, and planning is underway to reduce these differences over time to facilitate trade. State legislators must also consider the fact that if they set State excises substantially

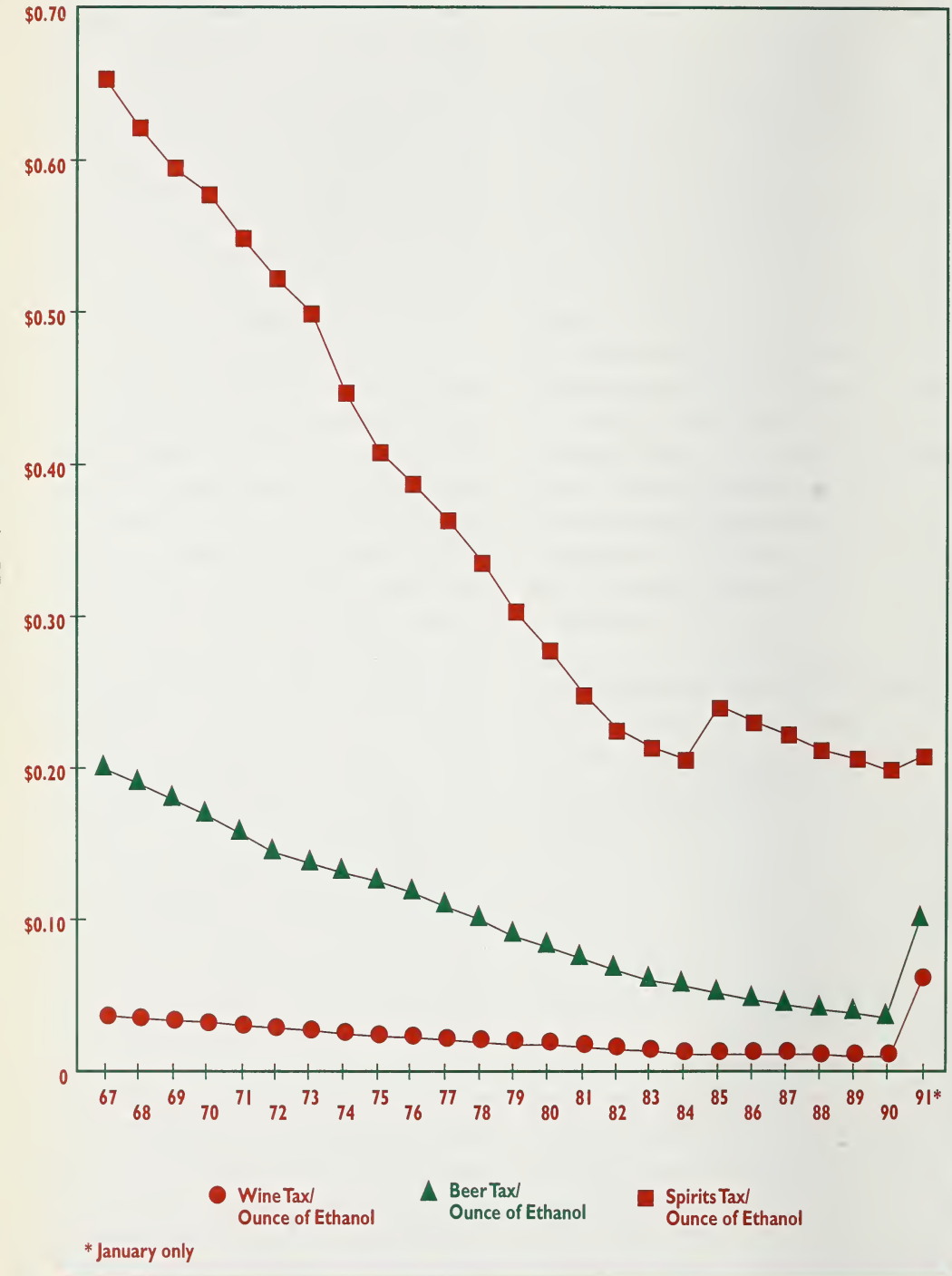
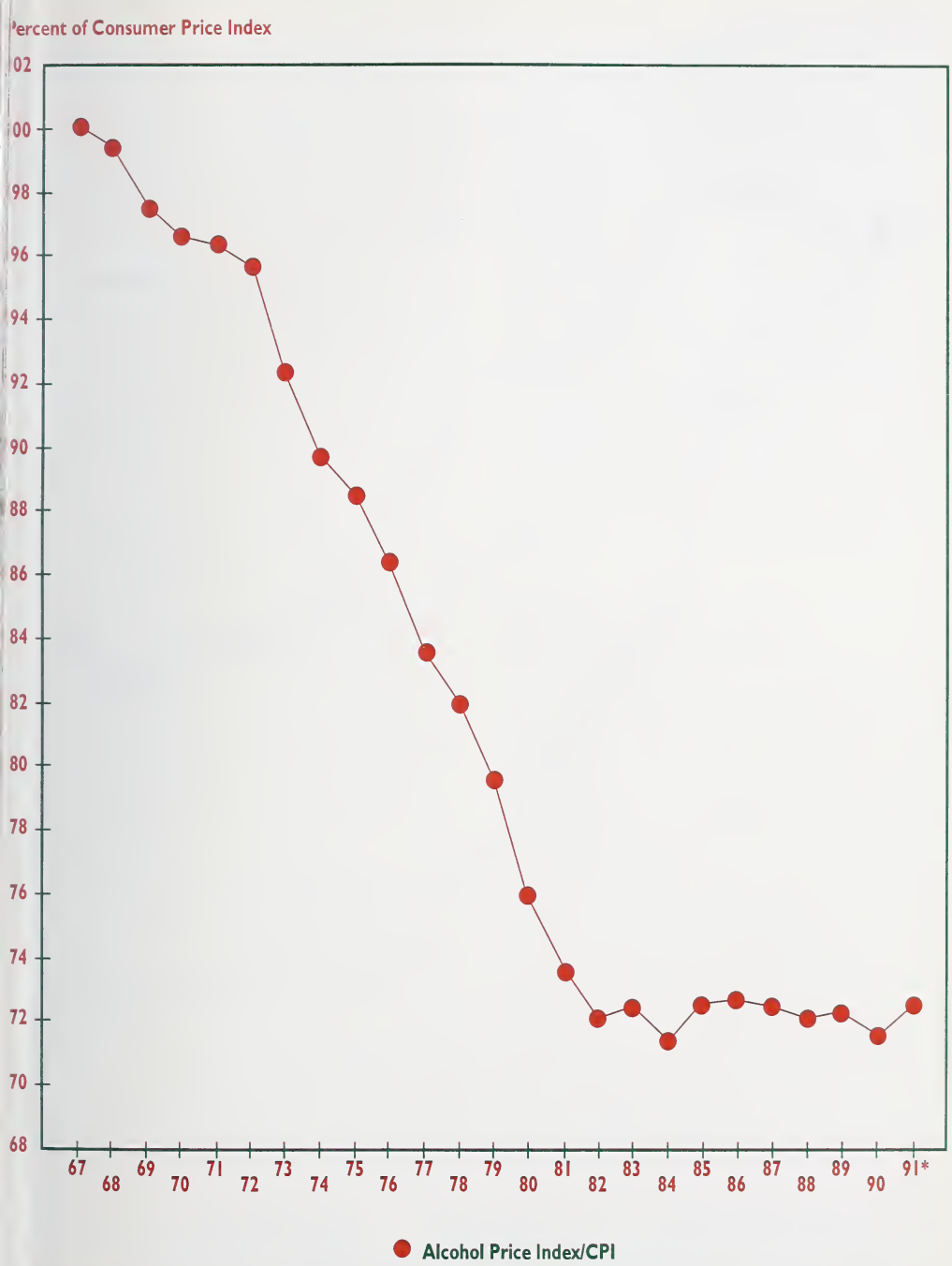


FIGURE 1: *Federal Excise Taxes in Constant (Jan. 1991) Dollars, 1967–91*



* January only

FIGURE 2: Average Price of Alcoholic Beverages (Relative to CPI), 1967–91

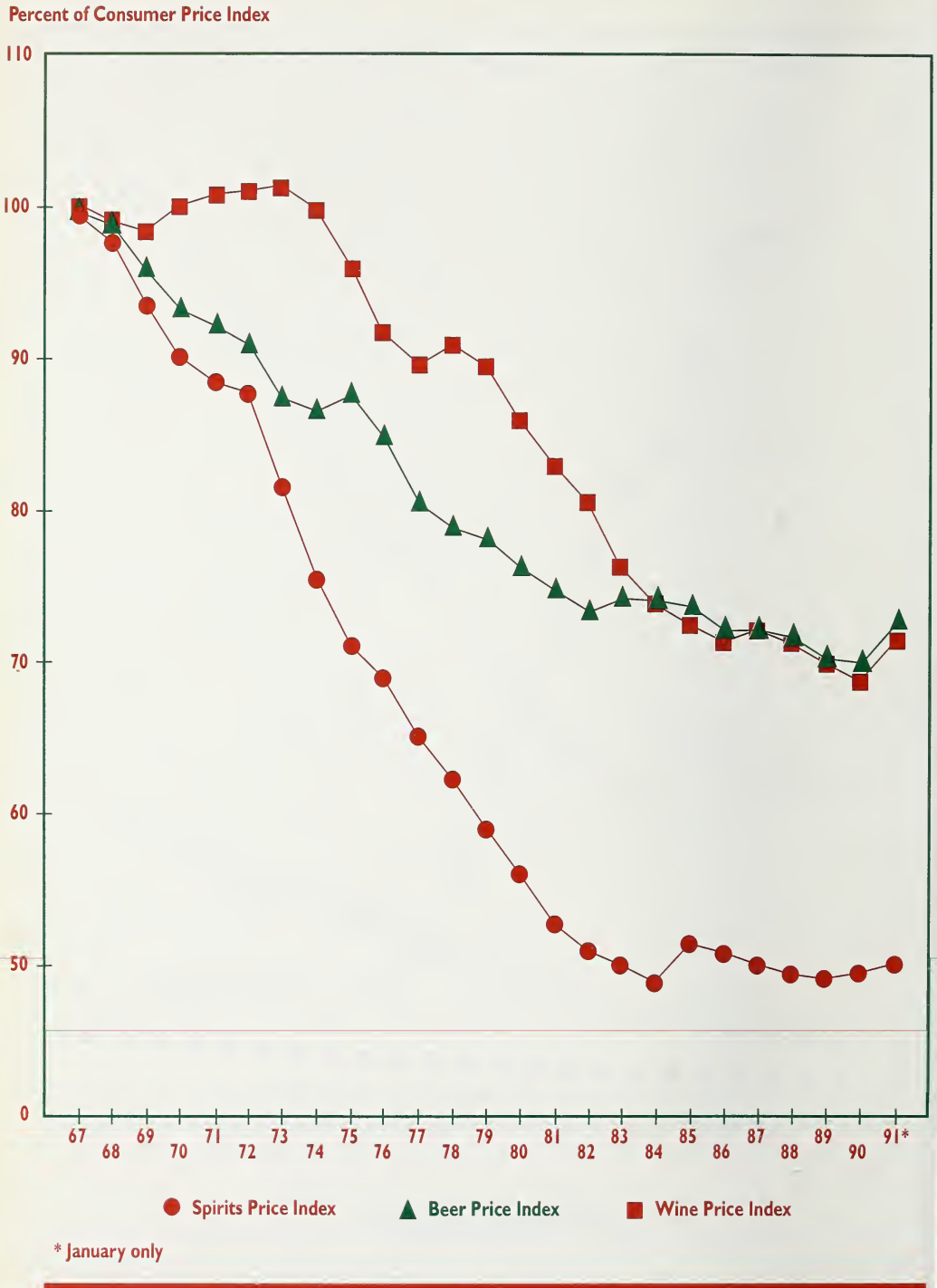


FIGURE 3: Average Prices of Beer, Wine, and Spirits (Relative to CPI), 1967–91

higher than those in nearby States, their residents will make a larger percentage of their alcohol purchases out-of-State. But these practical considerations aside, cross-jurisdictional comparisons do not seem to have much normative force in the public debate.

THE FAIRNESS STANDARD

Standard treatments of normative tax theory focus on fairness as well as economic efficiency (Stiglitz 1988; Rosen 1988). The notion of fairness incorporates several criteria, none of which are easy to operationalize. Here are three such criteria:

- Equals should be treated equally (the "horizontal equity" criterion).
- Households with greater ability to pay should be taxed more heavily than households with less (the "vertical equity" criterion). Usually this criterion is strengthened to an assertion that taxes should increase (at least) in proportion to household income.
- Households that receive greater benefit from Government activities should be taxed more heavily than households that benefit less (the "benefit" criterion).

Both advocates and opponents of increased alcohol taxes make claims based on these criteria. The most fundamental issue is whether otherwise-similar households should be taxed differently because one purchases more alcohol than another (a possible violation of horizontal equity). If that basis for asserting inequality is accepted, there remain other questions of fairness. Opponents emphasize the alleged regressivity of alcohol taxes (a violation of the vertical equity criterion). Proponents assert that heavy drinkers tend to receive a disproportionate benefit from some Government programs (particularly social insurance, publicly financed health care and social welfare) and hence should pay more, according to the benefit criterion. Indeed, this "benefit" argument has been expanded in recent years to assert that drinkers should pay for all the external social costs of their drinking, not just those that involve Government activities. The alcohol tax, in this view, should be viewed as equivalent to a "user fee" or insurance premium.

Incidence and the Vertical Equity Criterion

However fairness is to be judged, it is relevant to determine the actual incidence of alcohol taxation. This task is more difficult than it may seem at first glance. Payments of alcohol excise taxes are made by wholesale dealers, but there is a reasonable presumption that they pass this cost along to retailers and, ultimately, the consumer, in the form of higher prices. An increase in the Federal excise tax on beer will result in the consumer paying more for a six-pack than previously, but it is not clear how much more. If the beer industry were fully competitive both in production and distribution, and there were no scale economies, then in the long run the price would increase by the amount of the tax. In fact, production is oligopolistic, with just two producers supplying 69 percent of the domestic market.[4] Therefore, theoretical results on competitive market behavior do not apply.

There are a number of theories of how prices are determined in oligopolistic markets; as Harvey Rosen (1988, p. 277) notes, "Depending upon your view of how oligopolists function, practically anything can happen when a tax is imposed." A related case involving the cigarette industry is worth recounting, since this industry is similar to the beer industry in that the two leading producers supply 70 percent of the domestic market.[5] In 1982, Congress raised the Federal tax on cigarettes by 8 cents a pack. Between 1981 and 1985, retail cigarette prices rose by 54 percent (30 percent in real terms), even though the Federal excise tax increase, combined with small increases in State taxes that occurred over the period, was only 16 percent of the 1981 average retail price (Sammartino 1990). One explanation for this disparity is that the tax increase assisted cigarette producers in reaching an implicit agreement to coordinate on large increases in prices. In any event, the price increases were far larger than the tax increase. Thus, this tax increase burdened consumers significantly more than the actual amount of the tax, while the owners of producing companies (the oligopolists) actually came out ahead. Tobacco growers were harmed by the reduction in cigarette sales induced by the price increases, which caused a corresponding reduction in the derived demand for leaf. An increase in alcohol taxes would affect the analogous groups, but the pattern of effects might be quite different, depending on how the producers respond.

There is some evidence concerning the typical markup on alcohol excise taxes. Cook (1981) found that increases in State taxes on spirits were marked up by 19 percent on the average from 1960 to 1975. A recent analysis of the effect of the increase in the Federal spirits tax of 1985 reported an average markup of about 30 percent.[6] Other analysts have offered widely differing estimates of the markup (CSPI 1989). Given the nature of oligopolistic rivalry, it is reasonable to suppose that markups do vary widely, depending on the circumstances of the industry when the tax increase is imposed.

Given that consumers pay 100 percent or more (possibly much more) of a tax increase in the form of increased prices, it is important in judging the fairness (defined in this case in terms of both horizontal and vertical equity) that we know how alcohol sales are distributed across income levels.[7] A first observation is

---

[4] *In 1990, Anheuser-Busch had a 45.9-percent share of the domestic market, while Miller had a 23.1-percent share. Three other brewers have shares in the 5- to 10-percent range. The source for these data is Standard & Poor's Industry Surveys: Food, Beverages, and Tobacco, Section 3, June 27, 1991, F28.*

[5] *In 1990, Philip Morris had a 42.0-percent share and RJ Reynolds had a 28.9-percent share of the domestic market in cigarettes. See Standard & Poor's Industry Surveys: Food, Beverages, and Tobacco, June 27, 1991, F34.*

[6] *The Executives' Alcoholic Beverage Newsletter International, February 10, 1989, reported that the direct result of the Federal tax increase was to add about 45 cents to the price of a typical fifth of spirits. The amount of the tax increase was $2.00 per proof gallon, or from 32 cents on a fifth of 80 proof up to 40 cents on a fifth of 100 proof.*

obvious: that households that are equal in terms of income are very unequal with respect to alcohol consumption. For the adult population as a whole, the top 10 percent of all drinkers (6.5 percent of all adults) who drink the most heavily account for fully half of all alcohol consumed (Malin et al. 1982). At the other end of the spectrum, survey evidence suggests that at least one-third of all adults abstain. Thus, the majority of adults drink on occasion, but sales and consumption are concentrated among a relatively few heavy drinkers. This same qualitative pattern is found at each income level.

If we take the average dollar sales of alcohol at each income level, we see a nearly proportional relationship. Based on the Consumer Expenditure Survey data for 1984–85, Sammartino (1990) found that the fraction of after-tax income spent by households on alcoholic beverages is about the same in the second, third, and fourth income quintiles (p. 31); this fraction is somewhat higher for the lowest quintile, and somewhat lower for the highest quintile, but the pattern in the broad middle range is close to proportional.[8] This pattern holds for beer, wine, and spirits, as well as for all three together. This result suggests that an *ad valorem* tax on alcohol would not be particularly regressive. However, alcohol excise taxes are defined on quantity rather than value, so Sammartino's finding is not directly applicable in judging whether the tax is regressive. It is possible that alcohol expenditures increase with income, due to increased quality rather than volume. Therefore, a final judgment concerning regressivity must await data on the relationship between income and the volume of (rather than expenditure on) household alcohol consumption.[9]

Remember, too, that the incidence of a tax increase is not determined solely by who buys the product, but also by how the tax increase affects producers and sellers. Rosen (1988, p. 266) makes the point with this example:

---

[7] *It should be noted that income is not the only basis on which fairness might be judged. Someone who believes that income is distributed unfairly would not be willing to judge the fairness of a tax simply on the basis of its relationship to income. Suppose, for example, that women are discriminated against in the labor market. Then, the fact that women drink less than men who have the same income might be seen as making this tax fairer than if the reverse were true; making the men pay a disproportionate share becomes a sort of correction for "fairness failures" in the labor market.*

[8] *If alcoholic beverage expenditures are measured relative to total expenditures, rather than after-tax income, then these expenditures tend to increase across the five quintiles. Lyon and Schwab (1991) offer a related finding that alcohol taxes are less regressive relative to lifetime income than contemporaneous income.*

[9] *Hilton (1988) reports somewhat relevant results from a 1984 national alcohol survey. He cross-tabulates five different measures of drinking with various demographic factors, including an undefined variable called "income." The measure of drinking that is most relevant in the current context is the percentage who report drinking 60 or more drinks per month. For men, this percentage jumps from about 18 percent for those with income under $30,000, to 27 percent in the $30–40,000 interval and 32 percent in the top category. Very few women of any income report drinking this much. Based on the findings for men, it appears that per capita alcohol consumption does increase with income, but it cannot be determined whether it increases more or less than in proportion to income.*

> Suppose that poor people spend a relatively large proportion of their incomes on wine, but that vineyards tend to be owned by the rich. Then on the uses of income side, the tax [on wine] redistributes income away from the poor, and on the sources side, it redistributes income away from the rich.

Judging the income-related fairness of the tax requires that both sides of the ledger be taken into account.

A complete analysis of the consequences of a tax increase for producers requires that we consider how alcoholic beverage markets interact with other markets. So far, our discussion of the incidence of alcoholic beverage taxes has ignored possible effects of a change in alcohol taxes on other markets. The implicit assumption is that prices of other commodities and factors of production will not be affected by a tax-induced change in alcohol taxes. Because the alcoholic beverage industry is small relative to the size of the overall economy,[10] this assumption seems reasonable, but bears checking. A recent report on the effects of "sin" taxes (on both alcohol and tobacco) attempts to trace out secondary effects through use of a computable general equilibrium model (Boyd and Seldon 1991). The authors conclude that the distributional consequences of tax increases include losses to some asset owners, disproportionately those with low or high incomes. Those with incomes in the $20–40,000 range in 1984 suffered the smallest losses in proportion of their income.

It should also be noted that the consumption of alcoholic beverages has substantial external consequences for some other markets that are not the result of market transactions. For example, a tax increase on beer will reduce the traffic accident rate, which in turn will reduce the cost of driving, both directly and through reduced automobile insurance rates. A complete analysis of the incidence of the beer tax requires that we take account not only of who pays the tax, but also of who benefits from the reduced cost of driving. Depending on the strength of these external effects, it is possible that most households would gain as a result of a tax increase.

It is also reasonable to suppose that heavy drinking by the head of the household affects the welfare of dependents (Cook 1981). The head's expenditures on drinking may increase if his or her demand is price-inelastic; on the other hand, if a tax increase has the effect of reducing the head's drinking, the result may be to increase his or her productivity on the job and as a caretaker at home. It is likely, then, that dependents of some heavy drinkers benefit from a tax increase while others lose, depending on the resulting changes in expenditure and behavior of the head.

Finally, another feature of the alcoholic beverage market should be considered. Many drinks are consumed at

[10] *For the years 1984–87, consumer expenditures on alcoholic beverages averaged 1.2 percent of total consumer expenditures (Statistical Abstract of the United States, 1990, Table 713).*

someone else's expense. By one estimate, 20 percent of all alcoholic beverage sales are to businesses (Sammartino 1990, p. 76). Employees who are reimbursed for their drinking in connection with entertainment and travel, or who are entertained at office parties at their employers' expense, are receiving in-kind compensation that may be curtailed if there is a tax increase.

User Fee

In a May 1989 press conference, Surgeon General C. Everett Koop stated "I strongly recommend that States consider 5 cents a drink to be the minimum tax increase. After all, who could quarrel with a 'nickel a drink' user fee [increase] that will generate billions of dollars to help save lives?" (CSPI 1990). The notion of the alcohol tax as a user fee is related to the benefit principle of tax fairness. People who abuse alcohol benefit from certain Government programs more than nonabusers. For example, heavy drinkers have elevated morbidity, disability, and crime victimization, and hence claim a disproportionate share of Government expenditures on medical care, long-term care, and criminal justice services (Rice et al. 1990). Government revenues from alcohol taxes help defray these and other alcohol-related public expenses. It may seem only fair that the drinker should pay for the costly consequences of his drinking, and the alcohol excise tax is a mechanism for accomplishing this result.

Given that the bulk of alcohol-related costs are associated with rare events (most notably traffic accidents), this "user fee" is akin to an insurance premium. Insurance premiums differ among individuals, according to a schedule that reflects individual characteristics that are correlated with the likelihood of a claim. Alcohol taxes differ among individuals in direct relation to how much they drink, which is certainly a strong predictor of the likelihood of an alcohol-related problem (Moore and Gerstein 1981, p. 45). However, they do not discriminate with respect to any of a number of other correlates of alcohol-related problems, such as age, sex, prior history of alcohol-related problems, or drinking pattern. A 21-year-old man who drinks 7 beers a week in a single session and then attempts to drive home pays the same tax as a 40-year-old woman who drinks one beer with dinner every night. From the actuarial viewpoint, then, this tax is less than perfect. But overall, the incidence of this tax makes sense—the majority of the population who abstain or drink only rarely will pay little or no tax, while most of the tax liability will fall on the relatively small group who drink heavily and, as a result, account for a greatly disproportionate share of the costly consequences.

Heavy drinkers as a group benefit more than others from certain Government programs, including Government-financed medical care and disability payments through the Social Security system. Heavy drinkers also contribute less than others in the form of payroll taxes over their lifetimes, since a disproportionate number die young or retire early. A recent, comprehensive study of this matter (Manning et al. 1989,

1990) estimated these and other external costs associated with heavy drinking. Based on the results of this study, it appears that most of these costs are not financed by Government, but rather by private insurance companies or by innocent victims. They concluded that the present value (using a 5-percent discount rate) of external costs per ounce of ethanol consumed was about 48 cents, double the average State and Federal tax per ounce, which suggests that the current "user fee" is not nearly high enough. However, this judgment is based on a more inclusive standard than the benefit criterion, which is limited to benefits provided by Government. The notion here is that the drinker should pay *regardless* of whether the cost is to a Government program or to a private organization or individual. While this expansion of the benefit criterion may seem reasonable, it lacks the reimbursement feature of the traditional criterion. The taxes paid to Government will not be transferred to those who actually bear the costs.[11]

The analysis by Manning et al. is unique in attempting to sort out alcohol-related costs between those that are borne by the drinker or his household (internal) and those that are borne by those outside his household (external).[12] By this reckoning, earnings lost as a result of heavy drinking are internal costs, whereas payroll taxes not paid as a result of early death or retirement are external costs. Medical and disability costs that are reimbursed by insurance are internal if the insurance premium is paid by the drinker and reflects his true risk status, and external otherwise. Traffic fatalities and injuries are also divided between internal (where the injury is to the drunk driver himself) and external (where someone else is injured in an accident involving a drunk driver). These and other accounting rules are not uncontroversial (Cook 1991; Cordes et al. 1990), but the intention is valid.

Manning et al. express their conclusions in terms of the external cost per "excess ounce," a term that requires definition. Some of their results are based on survey data where self-reported drinking can be matched up against medical expenditures and lost work; they find that those who report drinking more than two drinks each day on the average incur costs greater than those who drink less. In that sense, self-reported drinking above two

[11] *If the alcohol tax is seen as a user fee, then it does seem reasonable that its revenues be dedicated to compensating for alcohol-related costs, much as the Federal excise tax on gasoline is dedicated to the Highway Trust Fund. Some States have dedicated all or part of their alcohol tax collections to financing alcoholism treatment and prevention (CSPI 1990).*

[12] *The question of where the line should be drawn between internal and external is especially troubling with respect to intrafamily effects. Family members have individual interests that are sometimes in conflict, but there is a reasonable expectation that these differences can be negotiated more easily than is true for unrelated individuals. Nonetheless, there is a clear public stake in preventing alcohol-induced family violence, child abuse and neglect, and fetal damage. More problematic is the case where heavy drinking by the household head reduces his dependents' standard of living. Interestingly, in the 19th century, "dram-shop" liability provided the wife of an alcoholic a course of action against the saloon where her husband was drinking away the household means of sustenance.*

drinks each day is defined as "excess."[13] Data on fires, crime, and traffic accidents come from sources that provide some information on whether drinking was involved at the time, but no information on average consumption. The researchers assumed that the drinkers in these cases would report (if asked on a survey) an average consumption of greater than two drinks per day, and hence, the costs associated with these events can be attributed to "excess" ounces of alcohol.

> Each actual excess ounce of alcohol leads to: a 26 cent increase in undiscounted external medical costs; a 6 cent rise in covered work loss; small effects on group life, pensions and nursing home costs; 35 cents less in taxes on earnings; 58 cents for lost lives of innocent bystanders; and 35 cents for fire, crime, and other property damage; for a total external cost of $1.56 per excess ounce (Manning et al. 1990, p. 127).

When the delayed costs of current consumption are discounted to the present, using a 5-percent discount rate, the external cost is reduced to $1.19 per excess ounce. For the purpose of calculating an appropriate tax, they note that it is more appropriate to divide total external costs of drinking by all ounces for all drinkers, instead of excess ounces for heavy drinkers. That calculation yields the estimate of 48 cents per ounce (p. 127), well above the average excise tax per ounce, which they estimate to be 23 cents as of 1985.

There are other prominent estimates of the social cost of alcohol abuse, most recently that of Rice et al. (1990). Manning's estimated total cost of about $25 billion[14] is considerably lower than Rice's estimated total of $70 billion. A major difference is that Manning does not include lost earnings in his estimates, while Rice does. Under the "drinker should pay" criterion, Manning's estimate is more relevant because internal costs are paid by the drinker by definition.

Finally, we note the interesting ambiguity concerning the "drinker should pay" benefit criteria. Suppose that the alcohol excise taxes were increased so that total collections were equal to the external costs associated with drinking. The increase in these tax rates would cause a reduction in tax collections from other sources. As Sammartino (1990, p. 76) explains,

> Most revenue estimates for Congressional legislation make the standard assumption that nominal gross national product is fixed, and the tax increase leads to a reduction in factor incomes. The reduction in nominal factor incomes leads, in turn, to a

[13] *Manning et al. note that drinkers tend to underreport their consumption of alcohol by about 60 percent, so that those who report two drinks per day may in fact average about five drinks per day.*

[14] *The breakdown of social costs in the Manning et al. (1989) estimates in billions of dollars (1986) is as follows: $12.3, value of nondrinkers' lives lost in traffic accidents; $5.55, net medical and pension costs (5-percent annual discount rate); $3.6, property damage from traffic accidents; $3.1, criminal justice costs; and $0.6, fires and social program costs.*

decline in both income and payroll taxes. The revenue yield is usually assumed to be only 75 percent of the gross tax increase.[15]

The question, to which we have no answer, is whether fairness requires an excise tax increase large enough to match revenue from that tax to the external cost or, alternatively, does it require a still larger tax increase to compensate for revenues lost from other taxes.

ALCOHOL TAXES AS A PUBLIC HEALTH MEASURE

The previous section developed the argument in favor of taxing alcoholic beverages at a higher rate than other commodities on the grounds that it was only fair that the drinker should pay for the external costs of alcohol abuse. Alternatively, it can be argued that high taxes are warranted because they curtail alcohol abuse. Indeed, there is substantial evidence that a tax increase is directly beneficial to the public health. To the extent that reducing morbidity and mortality rates is in the public interest, this evidence is justification for raising rates. An economist would qualify the public health goal to distinguish between self-inflicted harm and harm inflicted by drinkers on others. We will discuss that distinction in the next section.

As is true for other commodities, aggregate consumption of alcoholic beverages is sensitive to price. Less obvious is whether a general price increase for alcoholic beverages will reduce consumption of one component of these commodities, ethanol. In the face of higher prices for alcoholic beverages, consumers can make a variety of substitutions in an attempt to maintain their ethanol consumption without paying more; they can switch to cheaper brands, substitute at-home consumption for drinking at bars, or attempt to make their own ethanol. But in practice, these substitutions are not large enough to negate the price effect. The chapter in this monograph by Siu Fai Leung and Charles Phelps reviews the evidence, which clearly supports the notion that ethanol consumption falls in response to a general increase in the price of alcoholic beverages.

Given this result, the question remains whether alcohol abusers are responsive to an increase in price, or whether price-induced changes in aggregate consumption are limited to drinkers who are not abusers. It would be convenient for researchers in this area if there existed a fixed relationship between average consumption and the prevalence of alcohol abuse. Indeed, the well-known "single distribution theory," attributed to Sully Ledermann (Bruun et al. 1975), asserts the existence of a simple functional relationship between average consumption in a population and the prevalence of chronic heavy drinking in that population. But whether or not this controver-

[15] *Boyd and Seldon (1991), using a computable general equilibrium model, estimate that an increase in alcohol and tobacco taxes will increase net Government revenue by only 60 percent of the increase in collections on those taxes. The "loss" of 40 percent is the result of the tax consequences of the reallocation of economic activity induced by the alcohol and tobacco tax increases.*

sial assertion is valid, it is not enough for our purposes. As shown below, the major share of alcohol-related mortality is not due to the long-term effects of chronic heavy drinking, but rather the acute effects of inebriation.[16] A large share of such violent deaths involve people whose overall consumption level is in the moderate range (Moore and Gerstein 1981, pp. 42–47; Kreitman 1986). Indeed, there is no reliable statistical connection between the prevalence of inebriation and per capita consumption in a population (Skog 1991, p. 215).

The Centers for Disease Control (CDC) estimated alcohol-related mortality for 1987 to be 105,000, or 4.9 percent of all deaths in that year (CDC 1990). The important components of this total of alcohol-related deaths are as follows: 29 percent, accidents; 17 percent, homicide and suicide; 17 percent, diseases of the liver; 15 percent, cancer; 9 percent, stroke; 5 percent, alcohol dependence or psychoses; and 8 percent, all other. By this measure, then, nearly half (46 percent) of the alcohol-related mortality in the United States stems from accidents and intentional violence. By another conventional public health measure, "years of life lost before age 65," violent deaths constitute fully 80 percent of the "alcohol problem" (CDC 1990). If we take these results at face value,[17] they suggest that American public health officials should focus on acute episodes of abuse, since it is inebriation, rather than the effects of chronic heavy drinking, that causes accidents and violence.[18] By comparison, the chronic effects of excess consumption have a relatively small effect on life expectancy.

So within the public health perspective, the primary question becomes whether the alcohol-related fatal-injury rate is responsive to the price of alcoholic beverages. Several studies have reported affirmative evidence (Grossman 1989; Phelps 1988). Cook (1981) studied 39 instances in which States increased their liquor tax between 1960 and 1975, finding strong evidence that traffic fatalities in those States fell as a result. Coate and Grossman (1988), using survey data, found that the prevalence of frequent drinking by youths was highly responsive to State beer taxes; Saffer and Grossman (1987), using panel data on State traffic fatality rates, found a fairly large negative effect of the beer tax for each of the age groups 15–17, 18–20, and 21–24. While this evidence is persuasive, it is not complete; to date, there have been no studies

[16] Hilton (1987) reports that the frequency of heavy-drinking occasions is a strong predictor in survey data of alcohol-related problems.

[17] In fact, estimates of alcohol-related mortality are subject to considerable uncertainty. An expert panel of the Institute of Medicine (1987, p. 2a) noted the ". . .inability as yet to quantitate alcohol's share of specific morbidity for such problems as heart disease, oral cavity cancers, and accidents, each of which have multiple interacting causes."

[18] It should be noted, however, that suicide rates may be influenced by the prevalence of chronic heavy drinking as well as the incidence of intoxication. We know that suicide rates are very high among alcoholics (US DHHS 1987, p. 11). Chronic heavy drinking may contribute in various ways to the objective problems facing a depressed individual as well as reducing his emotional capacity to deal with these problems.

demonstrating a direct link between alcohol taxes and rates of violent crime or nontraffic accidents.

Similarly, the evidence of the link between alcohol prices and the prevalence of chronic heavy drinking is quite strong but incomplete. A pair of studies found that changes in State liquor taxes had a statistically discernible effect on the mortality rate from cirrhosis of the liver (Cook 1981; Cook and Tauchen 1982). Since alcohol-related liver cirrhosis is typically the result of many years of heavy drinking, this finding suggests that chronic heavy drinkers are quite responsive to the price of alcohol. There are several natural experiments that support this conclusion (Cook 1981). A more complete study that includes beer and wine taxes remains to be done.

One conclusion from this brief review is that increasing alcohol prices would save lives. From the public health perspective, therefore, it appears desirable to increase alcohol taxes, unless that would produce other consequences that are harmful to the public's health. Three possible consequences come to mind: substitution from alcohol to illicit drugs, the creation of a dangerous black market for alcohol, and decline in healthy drinking practices. Each of these are discussed briefly below.

First, it is possible that an alcohol tax increase will motivate some consumers to substitute illicit drugs for alcohol, and that as a result the net effect on the public health would be less advantageous than suggested above. Whether alcohol and illicit drugs are in fact substitutes for one another is unclear, and some evidence suggests that the opposite is true. Those who study adolescent drug use are impressed by the uniformity of the sequence with which youths move from one drug to another, with alcohol preceding marijuana and other illicit drugs (Kandel and Maloff 1983). If alcohol initiation was postponed (by higher taxes), then perhaps initiation of illicit drug use also would be postponed. This "gateway" theory suggests that in the developmental sequence, alcohol and illicit drugs are complements. This conjecture is supported by the fact that most drug abusers of any age are also drinkers.

The argument that alcohol is a *substitute* for illicit drugs rests on the belief that the fundamental goal of the alcohol and drug abuser is intoxication, with substitution of one intoxicant for another, depending on what is available. A recent study by DiNardo and Lemieux (1991) used data from interviews with high-school seniors over the course of 10 years (*Monitoring the Future*) to measure the effects of a change in the drinking age. The authors found that an increase in the legal age reduces the prevalence of alcohol consumption and increases the prevalence of marijuana use, suggesting that alcohol and marijuana are substitutes at the extensive margin. It is possible that alcohol and drugs are substitutes in the short run (as confirmed abusers seek an alternative means of getting high) and complements over the long run (as a result of the initiation sequence).

If there is substitution in the short run, what is the net effect on mortality rates? The findings (reported above) that

an increase in the beer tax reduces the youthful highway fatality rate demonstrates that even if there is substitution from beer to other sources of inebriation, the net effect is reduced prevalence of dangerous driving by youths.

The second problematic consequence of an increase in alcohol taxes is increased production of moonshine. Illicit production of distilled spirits was a nontrivial problem in the 1950's (when Federal liquor taxes were far higher than today) and has virtually disappeared in more recent times as the tax level has declined in real terms. When the moonshine business was thriving, there was reasonable concern about the health risks caused by the lead and other poisons introduced by the processes used by backwoods distillers, but at no time did this problem compare, from a public health perspective, with the injuries and deaths caused by drinking legal alcohol. If Federal taxes were increased very substantially, the moonshine industry might be revived, but it is hard to see this as a serious problem compared with the harms caused by inebriation.

Finally, there is the question of whether an increase in alcohol prices would result in a decline in healthy drinking practices. A growing body of evidence indicates that moderate drinking offers some protection against coronary heart disease. One theory is that alcohol elevates high density lipoprotein cholesterol (HDL-C) in the blood when taken in moderate quantities. Elevated HDL-C is, in turn, associated with reduced risk of heart attacks caused by arteriosclerosis. Several field studies have found that moderate drinkers, when compared with abstainers,[19] are indeed at reduced risk of heart disease (Ashley 1982; Rimm et al. 1991).

However, some uncertainty remains. It is possible that abstainers differ from moderate drinkers in ways that are relevant for heart-disease risk, but that have not been controlled for in the epidemiological studies. In the British Regional Heart Study, involving a large sample of middle-age men, it was found that, compared with drinkers, ex-drinkers included a higher proportion of men with ischemic heart disease, high blood pressure, or diabetes (Wannamethee and Shaper 1988). The best of the epidemiological studies, including the most recent (Rimm et al. 1991), attempt to eliminate this bias by excluding subjects from the sample who have a prior history of heart disease or who have reduced their drinking for health reasons. But there remains a real possibility that there are other, unobserved characteristics that are confounding the statistical results. We need to have a better understanding of the interaction

[19] *Rimm et al. (1991) found that the risk of coronary artery disease decreases as self-reported drinking increases over the complete range of drinking observed in their sample. Their data were taken from the Health Professionals Follow-up Study involving more than 50,000 men ages 40 to 75 years. Their highest drinking category is the open-ended category of over 50 grams of ethanol daily, and this category had the lowest relative risk of coronary disease.*

between personal health and drinking choices before being confident that the epidemiological evidence should be taken at face value (Shaper 1990).

Even if we accept the epidemiological evidence that drinking offers some protection against heart disease for people over age 40, there remains an open question about whether moderate drinking increases life expectancy for this group. Moderate drinking may increase mortality risks from certain other causes, including stroke (Lands and Zakhari 1991). Furthermore, epidemiological conclusions on these matters are not sufficient to assess the net effect of alcohol taxes on mortality; we must also know how the relevant groups adjust their drinking in response to price changes. Suppose that mortality increases monotonically with drinking below age 40 and has a U-shaped relationship to drinking after age 40. The net reduction in mortality from a tax increase then will depend on the relative price elasticities characterizing these two groups.

The bulk of the evidence supports the conclusion that increasing alcohol taxes would extend life expectancy. From the perspective of public health promotion, taxes should be higher than they are now. How high should they be raised? Presumably there is some point at which, all things considered, a further increase would cause more health problems (from illicit supplies, for example) than it prevents. But the existence of such a point of "negative returns" to increased taxes is purely conjecture.

ALCOHOL TAXES TO PROMOTE ECONOMIC EFFICIENCY

In the normative framework of public health, there is a high priority placed on reducing morbidity and mortality; if an increase in alcohol taxes helps achieve that purpose, then it should be done (Beauchamp 1980; Cook 1991). The usual normative framework in economics defines the public interest in terms of individual preferences, as revealed by choices in the marketplace and elsewhere. If people choose to place their own health and safety at risk by drinking too much, that choice is to be respected and is not in itself an argument for raising alcohol taxes. However, as discussed previously, not all of the costly consequences of alcohol abuse are borne by the drinker and his/her immediate household. If the price of alcoholic beverages does not properly reflect these external costs, then there will be a tendency for drinkers to ignore them in deciding how much to consume. Taxes on alcohol can then be justified as a mechanism for internalizing the external costs of alcohol abuse; ideally, the tax on a drink would equal the expected value of the external cost of consuming that drink. In practice, this justification overlaps with the public health argument, since the bulk of the external costs associated with alcohol abuse stem from injuries (Manning et al. 1989).

Taxes imposed as a correction for free-market prices that do not reflect full social costs are sometimes called "Pigovian" in honor of A.C. Pigou (1962), who developed the theoretical justification for such taxes. There is a fundamental difference between

the Pigovian principle and the principle discussed in a previous section that the "drinker should pay." Remember that "the drinker should pay" derives from a notion of *fairness*, the idea being that drinkers should pay enough to compensate those who suffer from the actions of alcohol abusers. The Pigovian principle, on the other hand, is justified in terms of efficiency. In the absence of the tax, some people will drink more than is efficient, in the sense that at the margin their drinks have less value (to them) than they cost the public. The Pigovian tax has the effect of raising the price to the "correct" level, reflecting the social costs of an additional drink, so that the drinker will limit his drinks appropriately. The total revenue generated by a Pigovian tax may well be greater than the total external cost, because it reflects the external cost of the marginal drink rather than of the average (inframarginal) drink.

It should be noted that the efficiency argument requires that corrective taxation alters consumers' behavior in the expected way. If an increase in the tax (and price) of alcoholic beverages had no effect on the prevalence of alcohol abuse, then it would not improve economic efficiency. Thus, the evidence that tax increases do reduce alcohol abuse, which was developed in the preceding section in support of the public health argument for special taxes, also helps justify the Pigovian argument for special taxes.

The application of the corrective tax principle to drinking has an obvious difficulty: unlike the tax rate, the external cost of a drink varies enormously depending on the personality and certain other characteristics of the drinker, the time and place of consumption, and how many drinks have been consumed already. It is possible to institute some crude differentiation in tax rates, but not enough. Peter Diamond (1974) demonstrated that when social costs differ with circumstances, then the value of that uniform tax which maximizes social welfare (under certain assumptions) is equal to a weighted average of the marginal external costs. Pogue and Sgontz (1989) have applied this theory to the case of alcoholic beverages.[20]

Pogue and Sgontz (1989) begin by postulating that there are two types of drinkers, nonabusers and abusers, and that the abusers impose external social costs that increase with the amount that they drink. In this case, the corrective tax is simply equal to a fraction of the marginal external social cost, where the fraction is equal to the abusers' portion of the total change in consumption induced by a change in price. The correct tax would increase with the number of abusers and the extent to which their drinking was more responsive to price than was true for nonabusers.[21] Understanding the intuition behind this result requires recognition of

---

[20] *The most thorough and formal account of the theory of optimal taxation of alcoholic beverages is presented in Barbara Weinstein's dissertation (1983).*

[21] *It is quite plausible that a 5-cent increase in the price of a drink would have a greater consumption effect on someone consuming 50 drinks per week than someone consuming five drinks per week. For example, if they had the same response proportionally, then the heavy drinker's reduction would be 10 times as great as the moderate drinker's reduction.*

the gains and losses associated with imposing a corrective tax. Society gains to the extent that abusers drink less, and hence, the total external cost that they generate is reduced. Society loses to the extent that nonabusers drink less, because their drinks have no external cost and have personal benefit greater than the untaxed price (which in the Diamond-Pogue-Sgontz model is assumed to equal the cost of production). The aggregate gain increases with the number of abusers and the extent to which their drinking is price-sensitive. The aggregate loss increases with the same two factors as they apply to nonabusers.

Pogue and Sgontz's model is obviously a stylized version of reality, but has the virtue that it can be used to generate empirical results. Based on their best guesses concerning the relevant parameters, they conclude that actual alcohol taxes in 1983 were about half of the optimal level.

A number of authors have suggested that the principle of corrective taxation be extended to account for the presumed fact that consumers tend to underestimate the internal costs of their drinking. If consumers tend to disregard certain costly consequences of their drinking out of ignorance or psychic myopia, then it is possible that a tax increase would bring their drinking closer into line with their "true" preferences (Atkinson and Meade 1974; Godfrey and Harrison 1990). The consequence is to increase the magnitude of the proper corrective tax (Pogue and Sgontz 1989).

This analysis has focused on economic inefficiency stemming from the fact that some consumers will drink more than is appropriate when they do not bear the full costs of their consumption decisions. Our implicit assumption has been that the price system functions well in all other respects and that commodity prices reflect marginal social costs. For any number of reasons, that is not true. One of the important reasons is that taxes on income have the effect of reducing the incentive to provide labor services on the market. Taxes reduce take-home pay and may distort such choices as how much to work, how much effort to expend while working, and how much to invest in education and training (Rosen 1980). One approach to correcting for the disincentive effects of income taxation is to impose special taxes on commodities that are complements to leisure and substitutes for investment in human capital (Kay and Keen 1986, p. 88; Slemrod 1990, p. 159; Corlett and Hague 1953). Clearly, alcohol is one such commodity. Indeed, it has long been believed that the ready availability of alcohol reduces effective labor supply in various ways—many business leaders supported Prohibition for just that reason (Cook 1991; Fisher 1926). So special alcohol taxes can be justified as a corrective device for incentive problems caused by other taxes, as well as a correction for costly externalities.

UNIFORM TAXATION

A can of beer, a glass of wine, and a shot of spirits all contain approximately the same amount of ethanol, but are taxed quite differently; the Federal excise tax on a shot of spirits exceeds the tax on a can of

beer by a factor of two and on a glass of wine by a factor of three. A number of commentators have advocated that tax rates be made uniform, the same per unit of ethanol, regardless of the type of beverage (CSPI 1990). If special taxes on alcoholic beverages are ultimately justified by the fact that such beverages are intoxicating, then it seems only sensible that the tax rate should reflect the concentration of the ingredient that causes the intoxication.

Do the tax principles discussed above provide any guidance on this matter? Historical precedent does not provide support for tax uniformity, because the practice of taxing spirits higher (per unit of ethanol) than other beverages has been a feature of American tax practice since the beginning. The principles of fairness and efficiency offer more support for uniformity.

In regard to fairness, it is sometimes asserted that beer should be taxed more lightly than spirits because it is the drink of the "working man." Be that as it may, the Congressional Budget Office study of consumer expenditures found little difference among the three beverage types with respect to their distribution across income categories. Expenditures on each of the beverage types tended to increase in proportion to income, at least over the broad middle range of the income distribution (Sammartino 1990, p. 31).

Whether a tax on beer is more or less regressive than a tax on spirits, there is a widely held belief that spirits are in some sense more intoxicating than beer or wine, and hence more subject to abuse, while beer is the "drink of moderation." If this is true, we could make the argument for different tax rates on the basis of the user fee principle. If spirits drinkers are more likely to drink to intoxication and to harm others, then a mixed drink warrants a higher "fee." However, much of the evidence works against this conclusion. Berger and Snortum (1985, p. 232) note that "The consistent observation is that most drinkers prefer beer, and that drinkers who prefer beer are more likely to drive while intoxicated (DWI) than those who prefer wine or distilled spirits". Furthermore, based on the results of a national survey, they conclude that beer is disproportionately preferred by higher risk groups: men far more than women, youths far more than older people, those who drink a lot during a typical session far more than those who drink moderately. While these results do not imply that tax uniformity is the best policy, they do suggest that uniformity is better than the current structure, where beer is taxed much lower than spirits.

Application of the efficiency principle leads to a similar conclusion, but by a longer argument.[22] Previously, we discussed the alcohol tax as a correction for the external costs of drinking, concluding that in the absence of such a tax, consumers will be induced to purchase more

[22] *One argument in favor of tax uniformity follows from the fact that the large differences in Federal excise taxes between wine and spirits have led producers to adopt inefficient recipes for producing cordials and liqueurs for which it is possible to substitute wine for spirits as one of the sources of ethanol in the mixture (GAO 1990).*

than is efficient. Similarly, if the structure of taxes across the three types of alcoholic beverage do not reflect differences in external costs, then consumers will be induced to purchase an inefficient mix of beverages. One way to understand this point is with an artificial example. Suppose that alcohol abusers only drank beer. Then, a uniform tax across beverage types would discourage consumption of wine and spirits even though (by assumption) these beverages contributed nothing to the external costs. It would be more efficient to tax only beer.

In general, the possibility of a nonuniform tax is desirable because we are pursuing conflicting objectives: reduce alcohol abuse (i.e., drinking that produces negative externalities) without penalizing moderate drinking. In this circumstance, a rule stipulating uniformity of tax rates across beverage types would generally be less efficient than a well-designed differentiated system of tax rates. But if the choice is between the current structure of taxes and a uniform structure of taxes, then it appears that uniform taxes are preferable.

Using the Pogue-Sgontz rule, the tax for each type of beverage should be equal to the product of two terms: the marginal external cost (per ounce of ethanol) caused by the abusers and the fraction of the reduction in consumption induced by a price increase that is accounted for by the abusers.[23] The parameters necessary to translate this rule into numbers are not known. However, if beer drinkers are more likely to abuse alcohol than are wine or spirits drinkers, as suggested above, then beer taxes should be higher per ounce of ethanol than other taxes—unless beer abusers tend to be less responsive to price (relative to nonabusers) than is true for the other beverage types. In the absence of clear evidence on these relative elasticities, we conclude that there is no justification in terms of economic efficiency for setting the beer tax lower than the liquor tax.

RESEARCH PRIORITIES

A high priority for future research is a complete analysis of the incidence of Federal and State excise taxes on alcoholic beverages. The Congressional Budget Office study (Sammartino 1990) was based on data from the Consumer Expenditure Survey, which recorded expenditures rather than volume of consumption. Surveys that include detailed questions on the volume of consumption often do not include adequate measures of after-tax income. There is one survey that has good data on both personal income and the volume of drinking (the National Longitudinal Survey of Youth), but its scope is limited to youths up through age 30.

A second difficulty with existing studies of incidence is that they have failed to address the oligopoly issue in considering how a tax increase affects prices. Assuming that the industry sets prices as if it were competitive seems unreasonable

[23] *In general, the rule will also depend on the cross-price elasticities, which in this discussion are assumed to be zero. For discussion of the empirical work on this matter, see Saffer (1989) and Saffer and Chaloupka (1991).*

given the very high degree of concentration in beer manufacture. Developing an empirical measure of the markup on tax increases is straightforward in principle for any given sample, if data on retail beverage prices are available. For example, Cook (1981) used panel data on annual average prices by State to estimate the markup on changes in State excise taxes. An appropriate estimation procedure is a fixed-effects regression model. A more sophisticated study would investigate the possibility that the markup will vary, depending on the industry structure at the wholesale and retail level.

A second priority for research is motivated by the growing evidence that moderate drinking reduces coronary heart disease risk in people over age 40. If this is true, then it becomes relevant to know how tax changes affect drinking at the extensive margin. If the decision of *whether* to drink is less sensitive to price than the decision of *how much* to drink, then a tax increase is less problematic (in its public health effects) than if the reverse pattern is true. The next step in this area would be to search for evidence linking alcohol availability to the prevalence of ischemic heart disease. We know that Prohibition reduced death rates from liver cirrhosis; did it also increase death rates from heart disease? Other natural experiments where alcohol availability changed dramatically could also be found. Another approach would be to explore the effects of State-level tax changes, using either State-level mortality rates (by analogy to Cook and Tauchen 1982) or micro panel data such as the National Health and Nutrition Examination Survey I and its followup.

Of course, it is just as important that we know how price changes affect the prevalence of heavy and abusive drinking. While there are some results on this subject (discussed above) much more remains to be done. Panel data on individuals (such as the National Longitudinal Survey of Youth) offer the greatest promise in studying changes in consumption levels. For rare events, such as alcohol-related birth defects, homicide victimization, and involvement in serious crime, it is probably necessary to use aggregate data.

There is an interesting set of questions on the degree to which different types of alcoholic beverages are substitutes for each other, and illicit drugs are a substitute for alcohol generally. On the first question we now have a natural experiment—the January 1991 increases in the Federal excise taxes on wine and beer were much larger than the increase in the tax on spirits. On the second question, the survey data from *Monitoring the Future* should be useful in determining whether interstate differences in beer taxes and the minimum legal drinking age affected the prevalence of marijuana use among youth. The preliminary results reported by DiNardo and Lemieux (1991), discussed above, are interesting, but do not exhaust the potential of these data.

A final priority concerns the effects of taxation on the long-term consequences of alcohol abuse. There is some evidence that heavy drinking by youths results in early termination of schooling and reduc-

tion in labor supply (Mullahy and Sindelar 1991), as well as a tendency toward alcohol abuse in later life. If true, then it is plausible that alcohol availability to youths has consequences over the life cycle in these dimensions. This matter can best be studied by use of longitudinal data (Cook and Moore, submitted).

ACKNOWLEDGMENTS

We benefited from the detailed comments on an earlier draft by Greg Bloss, Frank Chaloupka, and Mike Hilton and from the suggestions of Charles Clotfelter, Robert Conrad, William Gentry, Helen Ladd, and Jody Sindelar.

REFERENCES

Ashley, M.J. Alcohol consumption, ischemic heart disease and cerebrovascular disease. *J Stud Alcohol* 43(9):869–887, 1982.

Atkinson, A.B., and Meade, T.W. Methods and preliminary findings in assessing the economic and health consequences of smoking with particular reference to lung cancer. *J Royal Stat Soc Series A* 137:297–312, 1974.

Beauchamp, D. *Beyond Alcoholism: Alcohol and Public Health Policy*. Philadelphia: Temple University Press, 1980.

Berger, D.E., and Snortum, J.R. Alcoholic beverage preferences of drinking-driving violators. *J Stud Alcohol* 46(3):232–239, 1985.

Boyd, R., and Seldon, B.J. Changes in sin taxes: A general equilibrium perspective. *Land Econ* 67(3):365–374, 1991.

Bruun, K.; Edwards, G.; Lumio, M.; Makela, K.; Pan, L.; Popham, R.E.; Room, R.; Schmidt, W.; Skog, O-J; Sulkunen, P.; and Osterberg, E. *Alcohol Control Policies in Public Health Perspective*. Forssa, Finland: Aurasen Kirjapaino, 1975.

Center for Science in the Public Interest. *Impact of Alcohol Excise Tax Increases on Federal Revenues, Alcohol Consumption and Alcohol Problems*. Washington, DC: the Center, 1989.

Center for Science in the Public Interest. *State Alcohol Taxes: Case Studies of the Impact of Higher Excise Taxes in 14 States and the District of Columbia. Raising Revenues and Reducing Alcohol-Related Problems*. Washington, DC: the Center, 1990.

Centers for Disease Control. Alcohol-related mortality and years of potential life lost—United States, 1987. *Morbidity and Mortality Weekly Report* 39(11), 1990.

Coate, D., and Grossman, M. Effects of alcoholic beverage prices and legal drinking ages on youth alcohol use. *J Law Econ* 31:145–171, 1988.

Cook, P.J. The effect of liquor taxes on drinking, cirrhosis, and auto fatalities. In: Moore, M., and Gerstein, D., eds. *Alcohol and Public Policy: Beyond the Shadow of Prohibition*. Washington, DC: National Academy of Sciences, 1981. pp. 255–285.

Cook, P.J. The social costs of drinking. In: *The Expert Meeting on the Negative Social Consequences of Alcohol Abuse*. Oslo: Norwegian Ministry of Health and Social Affairs, 1991.

Cook, P.J., and Moore, M.J. "Drinking and Schooling." Submitted.

Cook, P.J., and Tauchen, G. The effect of liquor taxes on heavy drinking. *Bell J Econ* 13(2):379–390, 1982.

Cordes, J.J.; Nicholson, E.M.; and Sammartino, F.J. Raising revenue by taxing activities with social costs. *Natl Tax J* 43(3):343–356, 1990.

Corlett, W.J., and Hague, D.C. Complementarity and the excess burden of taxation. *Rev Econ Stud* 21:21–30, 1953.

Council of Economic Advisors. *Economic Report of the President*. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1987.

Diamond, P. Consumption externalities and imperfect corrective pricing. *Bell J Econ* 5:526–538, 1974.

DiNardo, J., and Lemieux, T. "Alcohol, Marijuana, and American Youth: The Unintended Consequences of Government Regulation." Working paper, University of California, Irvine, 1991.

Eusterbrock, C. Surviving the budget battle: U.S. Federal tax increases may hurt mid-tier brands. *Impactinternational* 5(23), 1990.

Fisher, I. *Prohibition at its Worst.* New York: The Macmillan Company, 1926.

Godfrey, C., and Harrison, L. Preventive health objectives and tax policy options. In: Maynard, A., and Tether, P., eds. *Preventing Alcohol and Tobacco Problems.* Vol. 1. Hants, England: Avebury, Gower Publishing Co., 1990. pp. 54–74.

Godfrey, J. Alcohol taxes: From the Whiskey Rebellion to the 1990 budget summit. *Tax Notes* 48(2):137, 1990.

Government Accounting Office. *Report to the Joint Committee on Taxation, U.S. Congress: Alcohol Excise Taxes: Simplifying Rates Can Enhance Economic and Administrative Efficiency.* GAO/GGD-90-123. Washington, DC: GAO, 1990.

Grossman, M. Health benefits of increases in alcohol and cigarette taxes. *Br J Addict* 84:1193–1204, 1989.

Hilton, M.E. Demographic characteristics and the frequency of heavy drinking as predictors of self-reported drinking problems. *Br J Addict* 82:913–925, 1987.

Hilton, M.E. The demographic distribution of drinking patterns in 1984. *Drug Alcohol Depend* 22:37–47, 1988.

Hu, T-Y. *The Liquor Tax in the United States, 1791–1947.* New York: Columbia University Press, 1950.

Kandel, D.B., and Maloff, D.R. Commonalities in drug use: A sociological perspective. In: Levison, P.K.; Gerstein, D.R.; and Maloff, D.R., eds. *Commonalities in Substance Abuse and Habitual Behavior.* Lexington, MA: Lexington Books, 1983. pp. 3–27.

Kay, J., and Keen, M. Alcohol and tobacco taxes: Criteria for harmonisation. In: Cnossen, S., ed. *Tax Coordination in the European Community.* Amsterdam: Kluwer Academic, 1986. pp. 85–111.

Kreitman, N. Alcohol dependence and the preventive paradox. *Br J Addict* 81:353–363, 1986.

Lands, W.E.M., and Zakhari, S. Alcohol and cardiovascular disease. *Alcohol Health Res World* 14:4, 1991.

Lyon, A.B., and Schwab, R.M. "Consumption Taxes in a Life-Cycle Framework: Are Sin Taxes Regressive?" NBER Working Paper No. 3932, Cambridge, MA, 1991.

Malin, H.; Coakley, J.; Kaelber, C.; Munch, N.; and Holland, W. An epidemiologic perspective on alcohol use and abuse in the United States. In: *Alcohol Consumption and Related Problems.* National Institute on Alcohol Abuse and Alcoholism Alcohol and Health Monograph No. 1, DHHS Pub. No. (ADM)82-1190. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1982.

Manning, W.G.; Keeler, E.; Newhouse, J.P.; Sloss, E.M.; and Wasserman, J. The taxes of sin: Do smokers and drinkers pay their way? *JAMA* 261(11):1604–1609, 1989.

Manning, W.G.; Keeler, E.; Newhouse, J.P.; Sloss, E.M.; and Wasserman, J. "Costs of Poor Health Habits." Unpublished manuscript. Santa Monica, CA: The Rand Corp., 1990.

Moore, M.H., and Gerstein, D.R., eds. *Alcohol and Public Policy: Beyond the Shadow of Prohibition.* Washington, DC: National Academy Press, 1981.

Mullahy, J., and Sindelar, J.L. "Alcoholism, Work, and Income Over the Life Cycle."

Cambridge, MA: National Bureau of Economic Research working paper 3909, 1991.

Phelps, C.E. Death and taxes: An opportunity for substitution. *J Health Econ* 7:1–24, 1988.

Pigou, A.C. *A Study in Public Finance.* 3rd ed. London: Macmillan & Co., Ltd., 1962.

Pogue, T.F., and Sgontz, L.G. Taxing to control social costs: The case of alcohol. *Am Econ Rev* 79:235–243, 1989.

Rice, D.P.; Kelman, S.; Miller, L.S.; and Dunmeyer, S. *The Economic Costs of Alcohol and Drug Abuse and Mental Illness: 1985.* Report submitted to the Office of Financing and Coverage Policy of the Alcohol, Drug Abuse, and Mental Health Admin., U.S. Dept. of Health and Human Services. San Francisco: Institute for Health & Aging, University of California, 1990.

Rimm, E.B.; Giovannucci, E.I.; Willett, W.C.; Colditz, G.A.; Ascherio, A.; Rosner, B.; and Stampfer, M.J. Prospective study of alcohol consumption and risk of coronary disease in men. *Lancet* 338:464–468, 1991.

Rosen, H.S. What is labor supply and do taxes affect it? *American Economic Review Papers and Proceedings* 70(2):171–176, 1980.

Rosen, H.S. *Public Finance.* 2nd ed. Homewood, IL: Irwin, 1988.

Saffer, H. "Alcohol Consumption and Tax Differentials Between Beer, Wine, and Spirits." NBER Working Paper No. 3200, Cambridge, MA, 1989.

Saffer, H., and Chaloupka, F. "Alcohol Tax Equalization and Social Costs." Unpublished manuscript, 1991.

Saffer, H., and Grossman, M. Beer taxes, the legal drinking age, and youth motor vehicle fatalities. *J Legal Stud* 16:351–374, 1987.

Sammartino, F. *Federal Taxation of Tobacco, Alcoholic Beverages, and Motor Fuels.* Congressional Budget Office, Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1990.

Shaper, A.G. Alcohol and mortality: A review of prospective studies. *Br J Addict* 85:837–847, 1990.

Skog, O-J. The prevention of alcoholism and drinking problems—Who are the target groups? In: *The Expert Meeting on the Negative Social Consequences of Alcohol Abuse.* Oslo, Norway: Norwegian Ministry of Health and Social Affairs, 1991.

Slemrod, J. Optimal taxation and optimal tax systems. *J Econ Perspect* 4(1):157–178, 1990.

Stiglitz, J.E. *Economics of the Public Sector.* 2nd ed. New York: W.W Norton & Co., 1988.

U.S. Department of Education, National Commission on Drug-Free Schools. *Toward a Drug-Free Generation: A Nation's Responsibility, Final Report.* Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1990.

U.S. Department of Health and Human Services. *Sixth Special Report to the U.S. Congress on Alcohol and Health.* DHHS Pub. No. (ADM)87–519. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1987.

U.S. Department of Health and Human Services. *Healthy People 2000: National Health Promotion and Disease Prevention Objectives.* Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1990.

Wannamethee, B., and Shaper, A.G. Men who do not drink: A report from the British Regional Heart Study. *Int J Epidemiol* 17:307–316, 1988.

Weinstein, B.G. "Optimal Liquor Regulation." Dissertation, The University of Michigan, Ann Arbor, 1983.

3

Alcohol Problems and the Control of Availability: Theoretical and Empirical Issues

Paul J. Gruenewald, Ph.D.

INTRODUCTION

There is a certain degree to which investigations of the relationships between access to alcohol and alcohol consumption and related problems appear to be discussions of the obvious. Few would doubt that sufficient increases in beverage prices, decreases in outlet densities, or elevations in the minimum drinking age could result in the reduction of consumption and, by definition, alcohol-related problems. While such dramatic alterations are unlikely to occur, in principle, changes in each of these controls on access to alcohol could lead to the elimination of alcohol consumption altogether. In practice, when such controls have been implemented, the results have been somewhat ambiguous. Thus, during Prohibition, alcohol-related problems appeared to both decrease (as measured by cirrhosis mortality rates) and increase (as measured by the illegal trade in alcohol and associated criminal activity) (Prendergast 1987). Although the experiment of Prohibition is widely held to have demonstrated the failure of this control measure, it is perhaps more informative to note that it demonstrated the resilience of the competing social values of the "good" of consuming alcohol versus the "good" of reducing alcohol-related problems. Alcohol control measures, the primary subject of discussion at this workshop, are simply the legal means through which these competing social values are politically realized. It is in this intellectual arena that research into alcohol control measures takes on economic, social, and political significance.

As will be argued in this review paper, the importance of studies of the effects of alcohol control policies upon alcohol consumption and alcohol-related problems is reflected in their ability to uncover the complex of economic, social, and political interactions that surround the purchase and consumption of alcoholic beverages. The sometimes dire consequences of drinking (e.g., deaths related to drunk driving) emerge in this context, motivated by the values individuals place upon the consumption of alcohol. While control policies attempt, first and foremost, to restrict

human activities in proscribed areas, they do not guarantee their own success. In a context of conflicting values, social forces operating to maintain the target behavior (drinking) may well circumvent the objectives of control. Thus, policies intended to reduce drunk driving crashes through decreases in outlet densities may result in increased traffic-related deaths as consumers maintain drinking levels and take on the additional risks of greater travel to alcohol outlets (Smart and Docherty 1976; Colon and Cutter 1983; Sprattler et al. in review).

The purpose of the current paper is to critically review the theoretical and empirical bases of our knowledge about the relationship of alcohol availability to alcohol consumption and alcohol-related problems. The literature on alcohol reflects the general view of researchers that restrictions on access to alcohol will result in reductions in alcohol use and alcohol-related problems. Access is taken to mean any of the broad social and economic mechanisms that enable the acquisition and consumption of alcohol, including both the social and economic environments of use, from membership in social groups that support drinking, to the marketplace for alcohol itself. In this context, the term availability is almost exclusively reserved to indicate the ease with which alcohol may be purchased for consumption either on or off premise (e.g., at restaurants and bars or at liquor and grocery stores). While availability in this sense can be construed to be affected by a far broader collection of regulatory activities than those to be discussed here (e.g., taxation and the setting of minimum drinking ages), the current focus will be upon licensing regulations influencing forms of distribution and hours and days of sale.

THEORETICAL FOUNDATIONS OF THE LITERATURE

Social Norms and the Distribution of Consumption

Historically, studies of the relationships between the availability of alcohol and alcohol consumption and related problems have relied upon a somewhat artificial theoretical dichotomy between two approaches to the social forces determining these outcomes: The social norm approach (represented well in the work of Frankel and Whitehead 1981) suggests that societal prescriptions for and proscriptions against the use of alcohol determine consumption patterns and, subsequently, problems. The distribution of consumption approach (represented well in the work of Bruun et al. 1975) suggests that, all else being equal, access to alcoholic beverages determines the mean consumption level in social groups and, subsequently, rates of heavy drinking and associated problems. These approaches can be seen as contrasting the driving forces behind consumption either as determined by societal expectations with regard to alcohol-related behaviors or as determined by access to alcoholic beverages. It can also be seen that the distribution of consumption approach lacks the necessary

motivational component for explaining the use of alcohol in the first place, a component itself attributable to social norms. Thus, the distinction between approaches is not a necessary one (Watts and Rabow 1983).

Necessary or not, the distinction between the approaches remains a basic theoretical reference point in the literature (Single 1988). This point, in general, is made to support the theoretical and empirical virtues of the distribution of consumption approach or one of its progeny. Although social norm theory will not be further discussed here, it should be noted that when surrogates for social norm measures have been introduced into well-defined empirical models of alcohol availability, consumption, and related problems, these measures have proven of substantive importance (Frankel and Whitehead 1981; Watts and Rabow 1983; Saffer and Grossman 1987; Gruenewald et al. 1992).[1] Despite its use as a straw man for theoretical argument, the social norm approach is little debated in the literature and would appear to have at least preliminary empirical support.

The distribution of consumption approach, on the other hand, assumes theoretical linkages between alcohol availability, its consumption, and related damage (Rush et al. 1986), which have been the subject of some debate. Restrictions on the availability of alcohol (through restrictions on outlet densities, for example) are assumed to alter the mean of the distribution of total alcohol consumption, which in turn alters the rate of heavy drinking and alcohol-related problems. The theoretical connection between the mean of the distribution of total alcohol consumption and heavy drinking is defended on the basis of the Ledermann (1956) model of the distribution of consumption, in which the first moment of this distribution determines heavy drinking rates. While this model is a good working hypothesis relating consumption to problems, and one that is partially validated in empirical research (see Cartwright et al. 1978; Cook and Tauchen 1982), it does not provide a necessary behavioral basis for a connection between aggregate consumption and problem rates (Duffy 1978; Tan et al. 1990; Lemmens et al. 1990). Thus, it is possible for increases in aggregate consumption to occur through alterations in drinking patterns unrelated to increased risk for alcohol-related problems (Cartwright et al. 1978; Mäkelä 1978). New individuals may begin drinking at low rates, or current drinkers may increase their total consumption in relatively safe ways (Gruenewald and Nephew 1991). With this point in mind, the distribution of consumption approach is sometimes reduced to the status of an empirically supported law of sorts. In this "law," consumption is related to problems in terms of an empirical

[1] *Regrettably, in a number of studies measures of the physical availability of alcohol have come to be confused with surrogates for social norms (Linsky et al. 1986). Thus, an ecological outcome of normative structures, the distribution of alcohol outlets, becomes confused with the normative structures themselves. In the studies cited, surrogates for social norms are clearly differentiated from measures of physical availability.*

model of availability, a model that assumes unspecified but direct relationships between availability, consumption, and problems (Single 1988; Mann and Anglin 1990).

Empirical models of availability, however, do not successfully avoid the problems reflected in the development of the distribution of consumption model. In the empirical models, a more defensible empirical formulation of the relationship between average consumption and problems is substituted for a relatively indefensible theoretical formulation of this relationship. Thus, as Skog (1985) has noted, the theoretical problem is not solved, but rather modified. He suggests that some model of the diffusion of normative behaviors with regard to alcohol may help explain why changes in mean consumption are accompanied by alterations in heavy drinking and alcohol-related problems. The current descriptions of this diffusion process are, however, too vague to warrant consideration as an explanatory theory of the relationship. An adequate theory of this process would identify what is diffused (total consumption, drinking patterns, etc.) and describe explicitly how the diffusion process operates.

Although not given voice in the literature, an alternative formulation of the relationship between the mean level of alcohol consumption and heavy consumption would be to assume that there are no necessary a priori relationships between these variables and that correlations between them arise through correlations with common exogenous determinants. In short, the attribution and search for a causal link between mean levels and heavy levels of consumption may be misguided. From this point of view, while there may be some diffusion processes acting to modify individual drinking behavior through social learning, and drinkers may migrate from lower to higher levels of consumption and back, it is the common causes of drinking (social norms, the psychosocial environment, ease of access for all drinkers, etc.) that induce correlations between rates of light and heavy drinking and, subsequently, rates of alcohol-related problems. The search for the common exogenous determinants of drinking patterns could lead to an explanation of the relationships between light and heavy drinking and specific risks for alcohol-related problems (Gruenewald and Nephew 1991).

An Economic Approach

Economic approaches to the evaluation of the relationship between the availability of alcohol and its consumption and related problems rest upon historically well-grounded theories of individual demand (Pindyck and Rubinfeld 1989; Godfrey 1989). The basic approach from this perspective is to assume that consumer demand for alcohol is determined by consumer tastes, alcohol beverage prices, personal disposable income, and the availability and prices of alternative commodities. The supply of alcohol is generally assumed to be infinitely elastic (Levy and Sheflin 1985), and the "full price" for alcoholic beverages is assumed to incorporate the additional costs of obtaining this commodity (Grossman 1988). Thus,

full prices, prices of alternative commodities, income, and measures of consumer "tastes" (such as are reflected in social norm surrogates), are generally assumed in the literature to be exogenous to demand. In most cases, demand itself is equated with total consumption measured in terms of aggregate beverage sales.[2]

With this set of assumptions in hand, and the technology of econometric estimation behind it, the economic approach to data on alcohol availability and consumption is to estimate elasticities of demand relative to these exogenous determinants. Assuming the broad applicability of the model to other behaviors, problem elasticities can be similarly estimated. Thus, the general economic theory of consumer demand is assumed to have specific application to the consumption of alcoholic beverages and to be generalizable to alcohol-related problems.

The virtue of this approach to the field is the provision of an explicit a priori model of the demand for alcohol with specific testable hypotheses regarding demand under different conditions. For example, since the full price of alcohol is assumed to include, among other things, the costs of obtaining this commodity, reductions in outlet densities are expected to reduce consumption by increasing the costs of travel to and from alcohol outlets. Similarly, social norms, or their surrogates, are assumed to determine tastes for alcohol, predisposing certain groups to greater or lesser levels of consumption. Further, based upon assumptions regarding the utility maximization of individual consumers, the general economic model of demand has very broad applicability. For example, if utilities for the consumption of alcohol differ among demographic groups, for whatever reasons, it can be expected that the demand functions for these groups will also be different.

The economic model of demand applied to alcohol availability, its consumption, and related problems, however, is not without its limitations. First, the assumption of the exogeneity of full prices in the demand model can be called into question. While one could argue that the supply of alcohol is infinitely elastic (given the ready availability of inputs and inexpensive costs of production), it is also arguable that the definition of full prices, including measures of availability such as outlet densities, leads to an endogeneity in the demand system. It has been suggested that the demand for alcoholic beverages places upward pressure on availability (Colon et al. 1982; Godfrey 1988; Gruenewald et al. 1992, 1993). Second, the economic model may be taken to assume that demand, measured in terms of total alcohol sales, reflects dimensions of alcohol consumption relevant to all consumers. It has, however, been noted that the consumption patterns of individual consumers (crudely, light versus heavy drinkers) may be related to their utilities for consumption and, hence, the price

[2] *That is, demand is not operationalized in terms of consumers' desires for commodities, but rather in terms of the degree to which these desires are met. This is a distinction of some importance in considering economic versus psychological explanations of the "demand" for alcohol.*

elasticities of their respective demand equations (Mäkelä 1978; Cook 1981). Third, the economic model as applied to alcohol problems inherits some of the difficulties suffered by sociological models based upon the distribution of consumption approach. As a simple empirical extension of the standard economic demand model, the application carries with it a rather superficial description of the relationship between consumption and problems. The empirical observation that a marginal increase in the cost of alcoholic beverages directly alters the rate of cirrhosis mortality among chronic alcoholics (Cook and Tauchen 1982) does not clarify the behavioral mechanisms underlying this relationship.

The mechanisms relating exogenous model terms (full prices) to problem outcomes remain inexplicit. From a policy point of view, the determination of these intervening mechanisms may not be essential to the use of economic models applied to problem data. After all, a stable marginal estimate of a price effect on problem rates is just that—a stable marginal estimate. However, from a research point of view, it is unsettling to see these gaps in the modeling process. The nature and causes of obtained elasticities remain to be explained. It is possible that within these explanations will reside suggestions for further preventive alterations in the socioeconomic system relating demand to problems.

A particularly good example of the current limits of our theoretical understanding of the relationships between consumption and problems arises in the explication of full prices for alcoholic beverages. The folding of measures of availability (hours and days of sale, outlet densities, etc.) into this economic concept (Grossman 1988) demonstrates the complexities of broadening the characterization of the demand for alcohol in this manner. Questions arise as to what aspects of availability are to be measured that are relevant to human behaviors with respect to alcohol and how to characterize these behaviors to come to a more complete definition of availability itself.

Subjective Availability

The larger sense in which availability may be construed has been discussed in the literature under the rubric of the subjective availability of alcohol (Smart 1980; Rabow et al. 1982; Neuman and Rabow 1986; Abbey et al. 1990*a*,*b*). From this theoretical perspective, availability is defined in terms of the perceptions of consumers who use alcohol. Subjective availability theorists contend that the physical availability of alcohol (measured, for example, in terms of outlet densities) serves as one of several inputs to a complete description of the impact of availability upon consumption and problems. Other (subjective) inputs include the willingness of individuals to pursue this commodity in their environment, the degree to which they consider prices in their purchasing decisions, and so on. On this basis it is asserted that physical availability, in conjunction with subjective availability, determines consumption. Thus, the causal determination of subjective availability by physical availability and con-

sumption by subjective availability is sometimes assumed to be demonstrated by empirically obtained correlations between measures of these constructs (Rabow et al. 1982). However, as noted by Abbey et al. (1990*b*), it is as likely that patterns of drinking may determine measures of subjective availability just as subjective availability may determine drinking.

While the theory attempts to separate physical from subjective availability (in terms of, for example, distances to outlets versus willingness to travel to them), it unintentionally blends conceptions of consumption patterns and the use of alcohol outlets (e.g., willingness to travel to an outlet may simply reflect one's drinking habits). This results in considerable causal confusion in attempting to explicate the basic notions of the theory. For example, it is difficult to establish to what degree the measure of willingness to travel to obtain alcohol represents aspects of physical availability (e.g., a high density of local outlets), versus differences in routine life activities that bring individuals into contact with alcohol (e.g., jobs located near customarily used outlets), versus differences in consumption patterns themselves (e.g., high frequencies of use). Each, and very likely all, of these features of the alcohol-purchasing environment may affect individuals' willingness to travel to obtain alcohol. The theory obscures rather than clarifies the relationships among these behavioral components of activities around alcohol.

Despite these problems, however, subjective availability theory has provided a new perspective from which to view the relationships of the availability of alcohol to alcohol consumption and related problems. The theory argues that perceptions of availability, in addition to physical availability, may well determine use. As the discussion offered here suggests, perceptions of availability themselves may be determined by physical availability as well as individual consumption and routine activity patterns. To the degree that the frequency and extent of drinking influence one's routine activities with respect to alcohol (purchasing alcohol as part of other regular activities, on special trips, on trips to and from work, etc.), and routine activities bring one into contact with alcohol (at restaurants, grocery stores, etc.), the physical availability of alcohol will be differentially utilized. In a sense, the full price of alcohol will in part be determined by consumption and routine activity patterns. In geographic areas with low densities of alcohol outlets, heavier drinkers may modify their routine activities in such a way as to maintain their contact with sources of alcohol (through purchases made, for example, in the context of other routine activities such as going to and from work). In geographic areas with identical outlet densities, consumption may be greater in that area in which alcohol is made available in conjunction with other routine activities (e.g., at restaurants).

The implication of these considerations for an economic model of the demand for alcohol is that the notion of a full price for alcoholic beverages will not be captured through the use of a simple

two-dimensional measure of the costs of alcohol to consumers expressed in terms of real prices and outlet densities. Rather, full prices will be sensitive to the nature of the routine activities in which individuals engage in their consumption of alcohol. While measures of the physical availability of alcohol can be included as exogenous determinants of the full price of alcohol in the standard demand model, the effects of changes in availability will be moderated by the routine activities, with respect to alcohol, engaged in by the subject population. Thus, the impact on changes in availability may be differentiated both quantitatively and qualitatively. Conditional upon the subject population, alterations in availability directly related to routine activities (such as in grocery stores) may have a far greater impact on consumption than equivalent alterations in availability relatively unrelated to routine activities (such as in liquor stores).

A Routine Activity Approach

Routine activity theory was originally developed as a means to explain aggregate rates of criminal victimization in terms of spatiotemporal features of human activities (Cohen and Felson 1979; Cohen et al. 1981). The theory, representing a human ecology of behavior, explains rates of victimization in terms of the availability of likely offenders and suitable targets and the absence of capable guardians against crime. The successes of the theory have been seen primarily in its ability to predict aggregate victimization rates in terms of a human geography of measures of activities that expose individuals to risk of victimization. For example, higher income groups have been found to have higher rates of personal larceny victimization because of enhanced exposure as victims (through travel, for example) and attractiveness as targets (Cohen et al. 1981). As noted by Cohen and Felson (1979), the theory has very broad applicability to any ecological model of behaviors related to routine human activities. They relate their model to fundamental functional human activities: gaining and losing sustenance, struggle for property, safety, territorial hegemony, sexual outlet, and so on. And they point out the fundamentally predatory relationship to be expected between criminals and their victims.

Since one of the characteristic human activities seen in much of the world involves the acquisition and use of alcohol, and since alcohol is often encountered in many of the activities of everyday life, it is reasonable to consider the applicability of routine activity theory to this aspect of human endeavor (see also Parker 1992; Treno et al. in press). The use of alcohol, of course, cannot be represented in terms of predator–prey relationships, as can models of victimization. Rather, use of alcohol is motivated by supportive biological, economic, social, and psychological conditions and entails the human activities required to obtain and consume this substance. Thus, routine activity theory calls for the identification of the primary spatial and temporal features of activities by which alcohol is obtained and consumed. Having identified these features of human activities related to alco-

hol, the entailed risks for alcohol problems can then be predicted.

A routine activities approach to the relationships between alcohol availability, its consumption, and related problems serves in several respects to clarify a number of issues discussed in this review. First, since the approach explicitly recognizes the contingent environmental nature of behaviors related to alcohol and their potential relationship to alcohol-related problems, emphasis is placed upon the development of ecologically valid models of the connection between consumption and problems. These models should include behaviorally relevant specifications of consumption patterns supplemental to the measures of total consumption customarily used in economic models (Gruenewald 1991; Gruenewald and Nephew 1991). Second, characteristic behaviors with respect to alcohol within social groups, often taken to be the basis of behavioral expectations regarded as social norms (Lewin 1951), are assumed to strongly determine both the nature and outcomes of preventive interventions. Third, the notion of a full price for alcoholic beverages characterized in the standard economic model receives support from a clear characterization of the relationships between physical availability, alcohol consumption patterns, and routine activities. Thus, the routine activities approach complements economic analyses of alcohol consumption.

As an example of the application of the routine activities approach to one aspect of alcohol consumption, consider models of beverage substitution addressed in the economics literature (Godfrey 1989). In models of the effects of price and outlet densities on alcohol consumption rates, it is customary to include effects for prices of close substitutes for the target beverage. For example, in measuring sales of spirits, increases or decreases in the price of a close substitute, such as wine, may result in increases or decreases in spirits consumption independent of spirits prices. Similarly, increases or decreases in spirits prices may result in increases or decreases in wine sales.

Measures of substitution in the literature on alcohol consumption have been solely based upon index prices for alternative beverages (beer, wine, or spirits). In addition to switching among beverage types, other forms of substitution are possible. Thus, within beverage types, the broad spectrum of available brands suggests that an increase in beverage prices may result in much substitution across brands. Additional opportunities for substitution arise, however, through alterations in routine activities associated with the purchase of alcohol. Thus, of the total variance in beverage prices observed in a study of pricing patterns in the State of California, more than 70 percent was attributable to differences among premise types alone (Treno et al. 1993).[3] That is, the greatest variance in beverage prices is related to where individuals buy this commodity (bars, restaurants, grocery stores, liquor stores, and so on). Obviously, the

[3] *Comparable figures for brand differences within beverage types and beverage type differences alone were 10 and 6 percent, respectively.*

greatest opportunities for substitution exist through alterations of routine activities with regard to the purchase of alcoholic beverages.

EMPIRICAL FOUNDATIONS OF THE LITERATURE

Naturally enough, the empirical approaches taken by researchers to the study of the relationships of the availability of alcohol to alcohol consumption and related problems reflect their theoretical predispositions. Social norm theorists, or those who consider social norms to be a significant part of the reasons people drink, include surrogates for social norms in their analytic models. Distribution of consumption theorists use some measures of alcohol availability directly related to consumption and problems. Economists, at the least, include measures of alcohol beverage prices, or taxes, and personal income as exogenous terms. The analytic models differ solely in terms of content, and the functions assumed to describe relationships among variables in the models are exclusively linear or estimably linear in form.

Unfortunately, the balkanization of theory in the area has been such that much research done from one perspective neglects the content of the next. In fact, the majority of studies reflect this theoretical myopia. Social norm theorists neglect economic variables (Frankel and Whitehead 1981), distribution of consumption theorists neglect social norm surrogates (Rush et al. 1986), and economists generally neglect availability measures (Levy and Sheflin 1985). While some research has begun to integrate these theoretical approaches (Watts and Rabow 1983; Godfrey 1988; Gruenewald et al. 1993), this review would be remiss without some strong encouragement to continue pursuing this line of work. With this in mind, however, it is important to acknowledge that a major problem with most studies of availability reviewed for this paper is their limitation to only a few of the most relevant variables for the study of alcohol consumption and related problems. In particular, the neglect of the primary economic measures of beverage prices and personal income renders most studies subject to criticisms of misspecification bias. This is a particularly telling problem since there is considerable evidence of substantial variation in beverage prices and incomes at State (Simon 1966*b*) and local levels (Treno et al. 1993) and of price and income effects (Godfrey 1989).

In addition to misspecification bias, many of the studies in the field suffer from a considerable variety of other methodological problems. These range from the complete absence of statistical tests supporting the assertions of the authors (Lenke 1982), to the use of enormous numbers of tests per degree of freedom (nine or more in Rabow and Watts 1982), through failures to control for nonstationarity in time-series designs (Smith 1987), to the selection of analyses based on their fit to prior theoretical assumptions (Ornstein and Hanssens 1985).[4] Most of these problems have been addressed in a variety of published articles. Of particular note here are the

methodological discussions of measurement issues using cross-sectional data in Rush and Gliksman (1986) and Rush et al. (1986), issues in the analysis of time-series data discussed in Wagenaar and Holder (1991), and an accessible review of issues in the analysis of time-series cross-sectional data by Hsiao (1986). These issues will not be further discussed here.

The approach of the current paper will be to briefly review the history of work and highlight the best studies available in the content areas of concern. Given the limitations of theory and method, and the sometimes very limited availability of data, the quality of these studies will vary considerably. In some areas the analyses presented by researchers are severely constrained by the availability of data. For example, analyses of the relationships between formal laws and access to alcohol founder on the almost exclusive availability of cross-sectional data. But these cross-sectional studies, nevertheless, represent our best guess to date of the relationships at issue. Other areas—for example, the study of the relationship of hours and days of sale to consumption and problems—appear to have adequate data but require further methodological development. Within these limitations, however, the current research provides a promising basis upon which to recommend future activity. For the interested reader, complete references to the studies reviewed for this paper are included at the end of this chapter.

Formal Regulation and Enforcement

Discussions in the literature of the relationships between formal alcohol control laws and alcohol consumption and problems date back to at least the time of Prohibition (Fosdick and Scott 1933). In the United States, empirical research on the issue seems to have begun in the 1960's with the work of Simon (1966*a*), who compared monopoly and license systems of alcoholic beverage control. Since that time a relatively small number of cross-sectional studies have been conducted comparing monopoly versus license systems in terms of their revenue-generating capacity (Zardkoohi and Sheer 1984; Swidler 1986) and in terms of their impacts upon alcohol consumption and problems (Smart 1977; Colon 1981, 1982; Gruenewald et al. 1992). Currently, all studies in this area suffer from a considerable lack of data. Single, relatively small ($n \leq 50$) cross-sections have been used in which comprehensive controls for exogenous determinants of unit differences have been either neglected (e.g., differences in beverage prices) or included to the detriment of the power of the statistical tests (Gruenewald et al. 1992; see Stevens 1986 for a discussion of this issue).

Interest in the monopoly–license distinction in the United States has contin-

[4] *Ornstein and Hanssens (1985) chose to report results of ordinary least-squares (OLS) analyses of time-series cross-sectional data based on the size of obtained price elasticities. Except under rather unusual conditions, OLS analyses of time-series cross-sectional data characteristically exhibit extreme bias (Hsiao 1986). The authors chose not to report alternative asymptotically consistent fixed effects models because of the small price elasticities obtained therein.*

ued unabated in the literature (Room 1987) despite the fact that empirical studies suggest that the primary remedial benefit related to alcohol consumption and problems to be accrued from a monopoly system is in the control on outlet distributions (Colon et al. 1982), and that the only essential difference between monopoly and license States is their control, specifically, of the distribution of retail outlets for spirits (Janes and Gruenewald 1991). Indeed, formal restrictions on the retail distribution of alcohol in monopoly States appear related only to ABC budgets (greater in monopoly States) and license densities (fewer spirits outlets, more wine outlets) (Gruenewald et al. 1992). To date, little is known about the relationship of the enforcement of ABC laws to alcohol consumption and problems. Gruenewald et al. (1992) do suggest, however, that enforcement appears greatest in those States with the strongest marketplace restrictions (e.g., restrictions on price competition).

The limitations on current research in this area are largely due to limitations on available data (see, for example, Kiholm Smith 1982). It is very difficult to obtain longitudinal data on alcohol beverage control agency activities, changes in the formal regulation of alcohol outlets, and so on. Therefore, much research is focused on the limited cross-sectional data available. Even with this restriction, it has been noted that the standard reference for formal laws regarding alcohol beverage control in the United States spans 10 volumes and some 15,000 pages of text (Janes and Gruenewald 1991). The historical tendency among researchers has been, therefore, to direct their efforts at understanding the effects of individual changes in formal law (e.g., changes in the minimum drinking age), or at the outputs of the control systems themselves (hours and days of sale, outlet densities, and so on), rather than tackle these more general aspects of the alcohol control system (such as the organizational structure of control systems beyond the monopoly-license distinction).

Alcohol Outlets

Direct studies of the effects of the physical availability of alcohol upon consumption and problems are greater in number than those regarding formal regulation and enforcement; however, on the whole, they have also ignored the market variables of price and income as covariates. Examining cross-sectional State-level data, there is some evidence that measures of the density of alcohol outlets may be related to alcohol consumption and problem rates (Parker et al. 1978; Harford et al. 1979; Colon 1982; Colon et al. 1982). However, these studies suffer from serious problems in misspecification (neglecting sociodemographic and economic variables and possible sources of simultaneity bias) and statistical testing (reliance on multiple testing procedures). More recent studies have been executed at the county (Rush et al. 1986; Gliksman and Rush 1986) and city (Watts and Rabow 1983) levels and are of importance for their efforts to avoid previous problems in statistically testing for the effects of availability. Using two different cross-sections (49

counties in Ontario, Canada, and 213 cities in California), these studies demonstrate stable and statistically significant relationships between outlet densities and sociodemographic background variables and consumption and problem rates (alcoholism, cirrhosis morbidity and mortality, arrests for public drunkenness, and traffic-related fatalities). Of these three studies, all neglect alcohol price effects in local markets, and the last neglects income effects upon alcohol consumption.

A critical subset of studies have included price and income in their analyses of time-series (McGuinness 1983; Walsh 1982; Godfrey 1988), cross-sectional (Schweitzer et al. 1983), and time-series cross-sectional (Ornstein and Hanssens 1985; McCornac and Filante 1984; Wilkinson 1987) data on availability and alcohol consumption. The first three studies suggest that availability measured in terms of per-capita outlet densities may be related to consumption rates but are limited by the unusually small numbers of available time-series observations (see Godfrey 1988). The fourth is a cross-sectional study of State data using a very limited subset of observed units (34 States) measured on a very large number of variables (up to 20). The fifth and sixth studies (Ornstein and Hanssens 1985; McCornac and Filante 1984) present ordinary least-squares analyses of time-series cross-sectional State data, a procedure that often produces biased estimates of effects using these kinds of data (Hsiao 1986). The seventh study (Wilkinson 1987) presents more appropriate generalized least-squares analyses of similar time-series cross-sectional data; the results suggest a small but significant relationship between per-capita outlet densities and alcohol consumption. This study, however, operationalized beverage prices through measures of tax revenue and sales data, creating a potentially biasing endogeneity between the measures of price and alcohol sales. These studies, each supporting the idea that availability is related to consumption, are also flawed because they do not take into account the potential simultaneity between these variables. That is, consumption and outlet densities may be related either by the former placing upward pressure on the latter or vice versa (Colon et al. 1982).

Godfrey (1988) has examined the issue of the endogeneity of availability and consumption in the context of time-series data similar to that used in McGuinness (1983) and Walsh (1982). She finds that there is some evidence that outlet densities are related to use (for spirits, wine, and beer) and that use is related to outlet densities (beer). Gruenewald et al. (1993), addressing this potential endogeneity using time-series cross sections of State data, show that outlet densities are related significantly to consumption, and that consumption places upward pressure, but not significantly, on alcohol outlet densities (for spirits and wine). Both of these studies include beverage prices and incomes as covariates. The latter also includes a subset of sociodemographic variables, believed to be related to consumption rates. This study, furthermore, is based upon a relatively substantial data base (114–288 time-series cross-sectional units).

One feature of the study by Gruenewald et al. (1993) that bears further discussion here is the particular definition of outlet densities that is used to predict alcohol sales. These authors formulate densities in terms of both population-based and geographically based metrics. Under the assumption that distributions of alcohol outlets track distributions of population gradients (that is, outlets are opened where populations of consumers are available to purchase this commodity), per-capita outlet densities reflect economic responses to aggregate population demand independent of the geographic environment of the market for alcohol. Under the assumption that populations are subject to different geographic spread due to differences in physical environment, land area per adult reflects geographic distances between people and, controlling for per-capita outlet densities, between alcohol outlets. Measuring both dimensions of the physical availability of alcohol, Gruenewald et al. (1993) show that both per-capita outlet densities and land area per adult are related to alcohol sales. Importantly, statistically controlling for variations in per-capita outlet densities, land area per adult is inversely related to sales. This finding strongly supports expectations based upon notions of a full price for alcoholic beverages.

A particularly relevant concern for more recent studies of availability is the estimation of elasticities of demand and problem elasticities relative to changes in outlet densities. If these effects, though significant, are insubstantial, then interventions aimed at reductions in availability (Wittman and Hilton 1987) may be of questionable value. Elasticities obtained using data from the United States can be derived from two studies: First, the cross-sectional study by Watts and Rabow (1983) provides estimates of problem elasticities relative to per-capita outlet densities. For public drunkenness arrest rates, on-premise densities had elasticities ranging from 1.7 to 2.1, and off-premise densities had elasticities of from 0.6 to 1.1. For cirrhosis mortality rates, on-premise densities had elasticities ranging from 0.3 to 0.7, and off-premise densities had elasticities ranging from 0.4 to 0.8. Second, Gruenewald et al. (1993) provide direct estimates of elasticities of per-capita outlet densities relative to sales of alcoholic beverages ranging from 0.03 to 0.4 for spirits and 0.3 to 0.4 for wine (effects for beer were not estimated). In terms of the physical geography of alcohol distribution (land area per adult), independent elasticities were estimated to be from –0.2 to –0.4 for spirits and –0.2 for wine. Thus, obtained elasticities are different from zero, and for some combinations of problems and forms of availability (drunkenness arrest rates and on-premise availability) elasticities may be greater than 1.0.

Although it is clear that some progress has been made toward the adequate determination of the effects of changes in outlet densities upon alcohol consumption and related problems, no study has yet been conducted to determine the relative costs and benefits of this approach to the prevention of alcohol-related problems. It is of particular

importance to determine the extent to which changes in physical availability simultaneously alter changes in consumption and problems. Because of the very high incidence of alcohol-related traffic crashes (Evans 1990) and the heavy dependence of U.S. citizens upon the automobile as their primary source of transportation, reductions in outlet densities may have a number of hidden costs, not the least of which is a potential increase in alcohol-related traffic fatalities due to increased driving exposure. Given striking differences in routine activities related to the purchase and consumption of alcohol at different outlets (e.g., the probability of driving after consuming alcohol when purchasing these beverages at restaurants versus liquor stores), the relative costs and benefits of reductions in outlet densities may vary strongly by outlet type.

Hours and Days of Sale

Except for their inclusion in several omnibus analyses of variables purportedly related to alcohol consumption rates (Hoadley et al. 1984; Ornstein and Hanssens 1985; Nelson 1988) and the inclusion of these variables in general measures of availability determined by formal laws (Smart 1977; Janes and Gruenewald 1991), studies of the effects of changes in hours and days of sale are notable in the literature for their general absence. The exceptions to this rule are a series of studies conducted by Smith (1987, 1988; see also Bibliography of Reviewed Materials) on a variety of changes in hours and days of sale made in various cities and states of Australia (see also Lind and Herbert 1982) and one descriptive study of the impact of extended operating hours at Scottish public houses and hotels (Bruce 1980). Although they contain a number of rather serious methodological flaws, these studies present some evidence, if only anecdotal, for impacts of changes in hours and days of sale upon a number of alcohol problems.

Smith (1988), for example, presents a study in which the introduction of Sunday alcohol sales in the city of Brisbane, Australia, is related to casualties and reported property damage from traffic crashes. Pre/post chi-square tests of problem rates aggregated over 2 years before and 3 years after the change in Sunday sales were used to test whether this change in availability had the expected impact. Estimates of the relative daily rates of problems were constructed by comparing Sunday rates to rates on other days of the week. Similar tests in surrounding comparison communities were used as controls for other possible contemporaneous changes affecting traffic crashes in the geographic area at large. Smith found that the measured relative daily rates did in fact increase in Brisbane, but not in the comparison communities, and that the time of day of these increases reflected the new opening hours of alcohol outlets. As he notes, the striking temporal relationships among these variables strongly suggests that hours and days of sale can have a substantial impact on alcohol problems.

There seems to be little doubt that, in this case, the change in hours and days of

sale in Brisbane was related to the daily distribution of problems on Sunday versus other days of the week. What is in doubt is precisely what this obtained relationship represents. The degree to which this result may be contaminated by other trend effects in Sunday sales, contemporaneous alterations in components of the alcohol control system (e.g., beverage prices or other aspects of availability), or other factors is unknown. And, more significantly, the degree to which the effect represents an increase in crash rates versus a redistribution of crash rates over time (days of the week) is not determined. Thus, on the basis of these data it is impossible to say whether the provision of Sunday sales had any aggregate impact on the problems of concern. Although the relative ratio of Sunday crashes to crashes on other days appeared to increase in response to a change in outlet opening days and hours, there is no evidence presented suggesting that this might not simply be a redistribution of the total rates of these events in time due to alterations in routine activities of consumers.

This particular empirical example from the field demonstrates the importance of improved evaluation in this area and the development of specific connections to issues regarding the routine activities of consumers. Given the available data on changes in hours and days of sale, and the available technologies for its data analysis, it would seem prudent to recommend a more thorough study of the relationship of this intervention to drinking and problem behaviors. Given the plausibility of an alteration in routine activities of consumers in response to new opening and closing hours of establishments, an exploration of the data with these particular activities in mind would seem particularly prudent.

Forms of Availability

In some respects each of the three areas of study previously discussed have dealt with a form of alcohol availability. Formal laws related to alcohol use always involve regulation of some aspect of the distribution or sale of these beverages, thus altering the form in which alcohol becomes available to consumers. Similarly, studies of outlet densities may consider the relative densities of on-premise versus off-premise establishments, and studies of hours and days of sale may consider differential impacts of redistributions of these variables across outlets. In the literature in general, however, studies of forms of availability have come to basically include studies of specific changes in alcohol control laws related to a redistribution of the physical circumstances through which alcohol becomes available to the consumer rather than quantitative changes in numbers of outlets or hours and days of sale, per se. Thus, studies of changes in the on-premise availability of alcohol may focus upon this change in form without emphasis upon quantitative changes in the numbers of outlets. While this distinction may have more to do with the form, to excuse the pun, than content of the particular studies, it will be followed here.

More than in any other area of research on the availability of alcohol, studies of the forms of availability have

exhibited an enormous range of theoretical and methodological quality. The range includes everything from anecdotal descriptions of time series developed in an effort to come to some a posteriori explanation of apparent effects (Williams 1975) to well-executed autoregressive moving-average (ARIMA) time-series models based upon a priori expectations of the impact of changes in the availability of liquor-by-the-drink (Holder and Blose 1987). In contrast to other areas of study, these studies have considered the impacts of alterations in alcohol control systems upon consumption and problems almost exclusively through the use of time-series data. Target alterations in the alcohol control system have included liquor store strikes (Mäkelä 1980), the provision of alcohol sales in grocery stores (Smart 1986), increases in off-premise availability (Mulford and Fitzgerald 1988), alterations in availability of liquor-by-the-drink (Holder and Blose 1987; Blose and Holder 1987*a,b*), and the privatization of liquor and wine outlets (Wagenaar and Holder 1991; Holder and Wagenaar 1991).

While a variety of relationships have been explored between the forms of availability and alcohol consumption and related problems, the methodological weaknesses of most of these studies have been such as to render interpretation of the outcomes ambiguous at best. These weaknesses include failures to account for nonstationarity in time-series data, the neglect of contemporaneous historical artifacts, and the exclusion of controls for the sometimes complex effects of time-ordered dependencies in endogenous model variables. Because these weaknesses are so common in analyses in this area, the current paper will focus upon two areas of work, studies of changes in laws regarding liquor-by-the-drink in North Carolina and studies of the effects of the recent privatization of spirits and wine sales in Iowa and West Virginia. These studies serve as exemplars of time-series work in the literature and the best serve to define the limits of this approach.

The work of Blose and Holder (1987*a*) demonstrates the application of ARIMA time-series modeling techniques to data on alcohol-related traffic crashes in North Carolina. The study was concerned with a change in liquor control laws that provided for sales of liquor-by-the-drink in counties that elected to enact this option. The researchers' interest was in determining the extent to which the provision of this new option for consuming beverages would effect an increase in alcohol-related traffic crashes measured on a monthly basis by the surrogate of single-vehicle nighttime crashes. Because some counties, did and some counties did not elect the liquor-by-the-drink option, a natural control for historical artifacts existed in the form of these no-change comparison counties. By comparing crash rates in change versus no-change counties, the authors could determine whether the effects of the intervention, if any, were an artifact of some unmeasured common historical effect on traffic crashes within the State.

The analyses of data from this study revealed a number of important points about evaluations of time series of this

kind. First, there were multiple time-ordered dependencies in the measured endogenous variable. These included significant monthly and seasonal trends and significant serial-correlated errors over several lags in each model. Thus, temporal dependencies in the data could not be represented by a simple linear trend, and serial correlations in errors exceeded the first-order autoregressive scheme most often assumed using linear regression techniques. Second, despite these complex time-ordered patterns in the data, it was possible to efficiently test for the effects of the intervention. Thus, the authors observed that there were uniformly significant increases in traffic crash rates in the change counties with no such increases in the no-change counties. Third, the provision of equivalent series from geographically contiguous and politically similarly defined areas (the no-change counties) provided an explicit control for potential historical artifacts of considerable face validity. Thus, although time-series analyses are always subject to potential misspecification biases due to historical artifacts, provision of an adequate comparison series can help deflect this criticism.

Wagenaar and Holder (1991) and Holder and Wagenaar (1991) have continued this line of work by applying ARIMA models to the analysis of the effects of the privatization of wine and spirits outlets on alcohol sales in the State of Iowa. In these analyses, ancillary controls were introduced for stocking effects (the rapid increases in beverage stocks at on-premise and off-premise establishments just before privatization) and national trends in alcohol sales. National trends were controlled through the inclusion of national rates of alcohol consumption as a correlate of State consumption over time. Analyses of adjacent State trends in consumption were used to determine whether historical biases arose during the time of the study. The analyses demonstrated substantive increases in alcohol sales related to privatization of outlets in Iowa. These effects were independent of rather strong stocking and national trend effects present in these data and, once again, no effects were found in adjacent States. The dependent measure, alcohol sales, demonstrated strong monthly and seasonal temporal trends and serial-correlated errors at multiple lags.

As a group these studies demonstrate the best execution of time-series studies in the literature to date. In particular, the modeling strategies applied to these data enabled the researchers to control for a variety of artifacts occurring naturally in the data (trend effects, serial-correlated errors, historical artifacts such as national trends in sales) and test for other sources of bias (specifically autoregression in the endogenous terms). The virtues of these time-series analyses of preventive interventions, however, are accompanied by a number of perplexing drawbacks. For example, in analyses of data on a change in a single State, it is impossible to be sure that other contemporaneous State changes might not be responsible for at least some of the effects observed. In this context, Holder and Wagenaar (1991) note that they were unable to obtain contemporaneous measurements of changes

in pricing and promotion activities. Nor did they measure local economic conditions or characteristics of consumers, such as personal income, which presumably changed during the same period. In addition, the changes in forms of availability studied do not represent simple univariate alterations in the alcohol control systems of each State. As noted by Holder and Wagenaar (1991), the privatization of liquor outlets in Iowa was accompanied by a growth in outlets and resulted in longer hours of sale and the opening of outlets over the weekend, permitted the use of personal bank checks and credit cards for the purchase of alcohol, and allowed advertising and price promotions. Since privatization in Iowa appeared to result in all these changes, from a research point of view it is difficult to attribute specific causes for the effects observed. Thus, one is left wondering about the degree to which increases in outlets, increases in hours and days of sale, decreases in beverage prices, and increases in advertising each contributed to the outcome observed.

The primary drawback with time-series studies, no matter how well executed, then, is the issue of attributed causes. The analysis techniques themselves (such as ARIMA models) are in many respects statements of ignorance. It is assumed that the temporal dependencies of the endogenous measures are generated by the series itself, although these dependencies could as well be generated by changes in exogenous factors. It is assumed that at the time of intervention only the change in intervention level itself is relevant to the change in the dependent measure, although other contemporaneous changes are likely. As noted by Holder and Wagenaar (1991), without replication across many States, or without many changes over time to examine, the impact of changes in forms of availability upon consumption and problems cannot be firmly established. In the words of an old German proverb, "If it happened once, it might as well not have happened at all." Wagenaar and Holder (1991) ameliorate this problem somewhat in their study of the effects of wine market privatization on consumption by studying changes in two States—Iowa and West Virginia—that took place at different times. The results of these analyses again show substantive effects of privatization on sales.

The skepticism about causal attribution aside, however, the presented studies would appear to provide some indication of the impact that major changes in alcohol control systems can have upon alcohol consumption and problems. Under the assumption that the model of the intervention process was appropriate and all other things being equal, the provision of liquor-by-the-drink in North Carolina did result in an increase in alcohol-related traffic crashes. Similarly, the privatization of wine and liquor outlets in Iowa and West Virginia resulted in some considerable increase in sales and, by inference, consumption and problems. With the further development of time-series data, and appropriate structural models for their analysis, it is likely that this approach to intervention evaluation will be a most powerful one.[5]

FUTURE DIRECTIONS

If the importance of studies in the area of alcohol availability is to be judged by their ability to uncover the complex of economic, social, and political interactions that take place around the purchase and consumption of alcoholic beverages, it is clear that we are just at the beginning of our work. While we are coming to know some of the ramifications of the regulation of alcohol outlets and hours and days of sale, and coming to appreciate the complexity of effects resulting from such large-scale changes as alterations in laws regarding liquor-by-the-drink and the privatization of alcohol outlets, little of our current knowledge bears upon the primary theoretical issues addressed in the introduction. Identifications of the specific relationships between alcohol sales, total alcohol consumption, alcohol consumption patterns, routine activities, and the incidence of individual alcohol problems remain to be developed. In this respect, the research described in this review demonstrates the efficacy of some changes in availability for altering consumption and problem rates. It does not inform us of the process by which changes in availability come to be realized in changes in these outcomes.

As suggested by the theoretical and empirical reviews, there are a number of fundamental problems that remain to be investigated in the area. In large part, these problems exist because the relationships between total alcohol consumption (often measured in terms of the more distal measure, sales) and alcohol-related problems have not been specifically defined. An adequate theoretical model of these relationships would explicate how individual consumption patterns, combined with individual routine activity patterns, come to be realized in individual risks for problem outcomes (Gruenewald and Nephew 1991). Measures of aggregate total alcohol consumption and aggregate problem rates reflect complex amalgams of these individual behaviors that may obscure some of the relationships we desire to investigate. Successful theoretical and empirical development of this research area will require complementary work at both the macro and micro levels.

Theoretical Issues

Foremost among the theoretical issues addressed in this review is the need to develop an adequate model of alcohol consumption patterns. A comprehensive model of drinking patterns, established at the individual level, is the only basis upon which theoretical connections to alcohol-related problems can be specifically derived. It is also the only basis upon which issues of aggregation bias can be theoretically evaluated. Limits upon the observed relationships between measures of total consumption and alcohol-related problems will be determined by the

[5] *Treno et al. (in press) have shown the applicability of multivariate time-series models to the evaluation of structural time-series relationships in an analysis of alcohol consumption rates in the United States. These techniques provide a method by which multiple time-series data can be effectively used to evaluate structural models.*

degree to which aggregate measures of drinking reflect aspects of individual consumption relevant to the genesis of specific alcohol-related problems. For example, measures of total consumption may be closely tied to the ultimate genesis of liver cirrhosis and only weakly tied to rates of drunk driving (Mann and Anglin 1990).

A second, but equally important, theoretical issue for the specific analysis of alcohol-related problems is the development of adequate models of routine activities. Except for medical outcomes directly associated with drinking, alcohol-related problems arise from the interaction of alcohol consumption with routine activities. Risks related to driving are amplified by the consumption of alcohol, as are risks for falls, drownings, and burns, the four leading sources of accidental death in the United States (Saltz et al. 1989). In the absence of routine activities related to these outcomes (e.g., driving, walking, swimming, and smoking), the incapacitating effects of alcohol have no effect on these mortality risks. In the presence of these activities, the incapacitating effects of alcohol strongly interact with problem rates (Gruenewald and Nephew 1991). Thus, a specification of routine activities related to alcohol use is essential for an adequate understanding of rates of associated problems. In addition, given the potential modifiability of routine activities with respect to alcohol (currently unknown), the plausibility of alterations in these behaviors for the prevention of problems remain to be explored.

A third, and supplementary, theoretical issue involves the need to explore further the applicability of the standard economic model of the demand for alcohol in the more complex environment of routine activities. As suggested in the theoretical review, routine activities may be modified to reduce the costs of purchase of alcoholic beverages for the purposes of maintaining consumption. In effect, individuals may "substitute" routine activities to reduce costs (Treno et al. 1993). In conjunction with the possibility that individuals may also alter consumption patterns to reduce risks related to alcohol consumption (Gruenewald and Nephew 1991), it is plausible to consider models of alcohol use in which individual consumers' demands for alcohol are constrained by the need to minimize costs of use through alterations in routine activities and alcohol consumption and purchasing patterns.

Empirical Issues

From the point of view of current research, and within the context of the theoretical developments of the current paper, it appears that continued empirical development along three fronts is called for.

First, the empirical review makes clear that much basic research in all four content areas remains to be performed: (1) The formal structure of regulatory agencies and the potential for utilizing their regulatory powers to prevent alcohol problems has been much discussed and little investigated. In particular, fluctuations in enforcement patterns would appear of particular relevance to the behaviors of outlets (with respect to sales to minors, violations of opening and clos-

ing hours, and so on). (2) Our understanding of the relationships of outlet densities to consumption and problems is at a nascent stage and requires, among other things, explorations of alternative measures of density (population versus geographically based), conclusive determination of the endogeneity of sales and physical availability, and explorations of the effects of density fluctuations on sales by outlet types. (3) The study of the effects of regulating hours and days of sale requires, as noted, a demonstration of the efficacy of this control measure in reducing alcohol consumption and an exploration of the potential for confounding alterations in routine activities with respect to daily consumption. (4) Studies of changes in forms of availability should continue to provide researchers with snapshots of the apparent impacts of large-scale changes upon alcohol consumption and related problems. Keeping the theoretical limitations of this approach in mind, such studies nevertheless do form the basis for motivating and directing future explorations of alcohol availability, its consumption, and related problems.

Second, the field requires essential, basic research to evaluate the costs and benefits of policies intended to reduce alcohol problems through alterations in availability. While we have come to know a small bit about the effects of rather large policy changes, and changes in outlet densities, upon alcohol consumption and problems, explorations of the social implications of these changes have not been made. For example, we do not know to what degree, if at all, reductions in outlet densities may lead to increases rather than decreases in drunk driving. Nor do we know whether reductions in densities of some outlet types (e.g., on-premise versus off-premise) may be more or less beneficial than others with regard to specific problem outcomes. Because the impact of a change in formal policy (e.g., the availability of Sunday sales) may have unexpected effects on drinking behaviors (e.g., no net change in consumption with a redistribution of drinking times), implementation of policies with inadequately explored outcomes can have unfortunate implications (increased drunk driving at high-risk times of the week).

Third, and finally, a great deal of empirical work also remains to be performed to secure an adequate foundation for theoretical developments of the relationships of alcohol consumption patterns to routine activities and the occurrence of alcohol-related problems. This work will help guide the theoretical development of micromodels of consumption and problems necessary to understand the outcomes of research performed using aggregate measures. Using survey research technology, it will be necessary to explore not only individual consumption patterns, but also when, where, and how individuals go about purchasing and consuming alcohol in their daily lives—the routine activities that surround consumption.

ACKNOWLEDGMENTS

I would like to acknowledge the many comments provided to me by a variety of readers of successive versions of this manuscript: Harold D. Holder, Robert N.

Parker, William R. Ponicki, and Andrew J. Treno. The comments of Drs. Louis Gliksman, Michael Hilton, and Gregory Bloss are particularly appreciated. After their careful review of the original manuscript, any faults in the presentation of the economic approach to the study of alcohol consumption and problems are necessarily my own.

Research and preparation of this paper were supported in part by National Institute on Alcohol Abuse and Alcoholism Research Center grant AA06282 to the Prevention Research Center and National Institute on Alcohol Abuse and Alcoholism grants 1-R01-AA08395-01 and 3-R01-AA08395-02 to the author, affiliated with Pacific Institute for Research and Evaluation.

REFERENCES

Abbey, A.; Scott, R.O.; Oliansky, D.M.; Quinn, B.; and Andreski, P.M. Subjective, social and physical availability. I. Their interrelationships. *Int J Addict* 25:889–910, 1990*a*.

Abbey, A.; Scott, R.O.; Oliansky, D.M.; Quinn, B.; and Andreski, P.M. Subjective, social and physical availability. II. Their simultaneous effects on alcohol consumption. *Int J Addict* 25:1011–1023, 1990*b*.

Blose, J., and Holder, H.D. Liquor-by-the-drink and alcohol-related traffic crashes: A natural experiment using time series analysis. *J Stud Alcohol* 48:52–60, 1987*a*.

Blose, J., and Holder, H.D. The public availability of distilled spirits: Structural and reported consumption changes associated with liquor-by-the-drink. *J Stud Alcohol* 48:371–379, 1987*b*.

Bruce, D. Changes in Scottish drinking habits and behaviour following the extension of permitted evening opening hours. *Health Bull* 38:133–137, 1980.

Bruun, K.; Edwards, G.; Lumio, M.; Mäkelä, K.; Pan, L.; Popham, R.E.; Room, R.; Schmidt, W.; Skog, O-J.; Sulkunen, P.; and Osterberg, E. *Alcohol Control Policies in Public Health Perspective.* Helsinki: Finnish Foundation for Alcohol Studies, 1975.

Cartwright, A.K.J.; Shaw, S.J.; and Spratley, T.A. The relationships between per capita consumption, drinking patterns and alcohol related problems in a population sample, 1965–1974. Part II: Implications for alcohol control policy. *Br J Addict* 73:247–258, 1978.

Cohen, L.E., and Felson, M. Social change and crime rate trends: A routine activity approach. *Am Sociol Rev* 44:588–608, 1979.

Cohen, L.E.; Kluegel, J.R.; and Land, K.C. Social inequality and predatory criminal victimization: An exposition and test of a formal theory. *Am Sociol Rev* 46:505–524, 1981.

Colon, I. Alcohol availability and cirrhosis mortality rates by gender and race. *Am J Public Health* 71:1325–1328, 1981.

Colon, I. The influence of state monopoly of alcohol distribution and the frequency of package stores on single motor vehicle fatalities. *Am J Drug Alcohol Abuse* 9:325–331, 1982.

Colon, I., and Cutter, H.S.G. Relationship of beer consumption and state alcohol and motor vehicle policies to fatal accidents. *J Safety Res* 14:83–89, 1983.

Colon, I.; Cutter, H.S.G.; and Jones, W.C. Prediction of alcoholism from alcohol availability, alcohol consumption and demographic data. *J Stud Alcohol* 43:1199–1213, 1982.

Cook, P.J. The effect of liquor taxes on drinking, cirrhosis, and auto accidents. In: Moore, M.H., and Gerstein, D.R., eds. *Alcohol and Public Policy: Beyond the Shadow of Prohibition.* Washington, DC: National Academy Press, 1981. pp. 255–285.

Cook, P.J., and Tauchen, G. The effect of liquor taxes on heavy drinking. *Bell J Econ* 13:379–390, 1982.

Duffy, J.C. Comment on "The single distribution theory of alcohol consumption." *J Studies Alcohol* 39:1648–1650, 1978.

Evans, L. The fraction of traffic fatalities attributable to alcohol. *Accident Analysis Prev* 22:587–602, 1990.

Fosdick, R.B., and Scott, A.L. *Toward Liquor Control.* New York: Harper and Brothers, 1933.

Frankel, B.G., and Whitehead, P.C. *Drinking and Damage: Theoretical Advances and Applications for Prevention.* New Brunswick, NJ: Rutgers Center for Alcohol Studies, Monograph 14, 1981.

Gliksman, L., and Rush, B.R. Alcohol availability, alcohol consumption and alcohol related damage. II. The role of sociodemographic factors. *J Stud Alcohol* 47:11–18, 1986.

Godfrey, C. Licensing and the demand for alcohol. *Appl Econ* 20:1541–1558, 1988.

Godfrey, C. Factors influencing the consumption of alcohol and tobacco: The use and abuse of economic models. *Br J Addict* 84:1123–1138, 1989.

Grossman, M. "Health Economics of Prevention of Alcohol-Related Problems." Paper presented at the workshop on Health Economics of Prevention and Treatment of Alcohol-Related Problems, National Institute on Alcohol Abuse and Alcoholism, Washington, DC, 1988.

Gruenewald, P.J. Loss of control drinking among first offender drunk drivers. *Alcohol Clin Exp Res* 15:634–639, 1991.

Gruenewald, P.J.; Madden, P.; and Janes, K. Alcohol availability and the formal power and resources of state alcohol beverage control agencies. *Alcohol Clin Exp Res* 16:591–597, 1992.

Gruenewald, P.J., and Nephew, T. "Consumption and Death: Dynamic Constraints on Patterns of Alcohol Use." Paper presented at the annual meeting of the Research Society on Alcoholism, Marco Island, FL, 1991.

Gruenewald, P.J.; Ponicki, W.R.; and Holder, H.D. The relationship of outlet densities to alcohol consumption: A time series cross-sectional analysis. *Alcohol Clin Exp Res* 17(1):38–47, 1993.

Harford, T.C.; Parker, D.A.; Pautler, C.; and Woltz, M. Relationship between the number of on-premise outlets and alcoholism. *J Stud Alcohol* 40:1053–1057, 1979.

Hoadley, J.F.; Fuchs, B.C.; and Holder, H.D. The effect of alcohol beverage restrictions on consumption: A 25-year longitudinal analysis. *Am J Drug Alcohol Abuse* 10:375–401, 1984.

Holder, H.D., and Blose, J. Impact of changes in distilled spirits availability on apparent consumption: A time series analysis of liquor-by-the-drink. *Br J Addict* 82:623–631, 1987.

Holder, H.D., and Wagenaar, A.C. Effects of the elimination of a state monopoly on distilled spirits' retail sales: A time series analysis of Iowa. *Br J Addict* 85:1615–1625, 1991.

Hsiao, C. *Analysis of Panel Data.* New York: Cambridge University Press, 1986.

Janes, K., and Gruenewald, P.J. The role of formal law in alcohol control systems: A comparison among states. *Am J Drug Alcohol Abuse* 17:199–214, 1991.

Kiholm Smith, J. An analysis of state regulations governing liquor store licenses. *J Law Econ* 25:301–319, 1982.

Ledermann, S. *Alcool, alcoolisme, alcoolisation, Vol. 1. Donnees scientifiques de caractere physiologique economique et social.* Institute National d'Etudes Demographiques, Travaux et Documents, Cah. No. 29. Paris: Presses Universitaires de France, 1956.

Lemmens, P.H.H.M.; Tan, E.S.; and Knibbe, R.A. Comparing distributions of alcohol consumption: Empirical probability plots. *Br J Addict* 85:751–758, 1990.

Lenke, L. Alcohol and crimes of violence: A causal analysis. *Contemp Drug Prob* Fall:355–365, 1982.

Levy, D., and Sheflin, N. The demand for alcoholic beverages: An aggregate time-series analysis. *J Public Policy Marketing* 4:47–54, 1985.

Lewin, K. *Field Theory in Social Science.* New York: Harper and Brothers, 1951.

Lind, B., and Herbert, D.C. The effect of "Sunday trading" on traffic crashes. Traffic Accident Research Unit, Traffic Authority of New South Wales, SR 82/112, 1982.

Linsky, A.S.; Colby, J.P.; and Straus, M.A. Drinking norms and alcohol-related problems in the United States. *J Stud Alcohol* 47:384–393, 1986.

Mäkelä, K. Level of consumption and social consequences of drinking. In: Israel, Y.; Glaser, F.B.; and Kalant, H., eds. *Research Advances in Alcohol and Drug Problems.* Vol. 4. New York: Plenum Press, 1978. pp. 303–348.

Mäkelä, K. Differential effects of restricting the supply of alcohol: Studies of a strike in Finnish liquor stores. *J Drug Issues* 10:131–144, 1980.

Mann, R.E., and Anglin, L. Alcohol availability, per capita consumption, and the alcohol-crash problem. In: Wilson, R.J., and Mann, R.E., eds. *Drinking and Driving: Advances in Research and Prevention.* New York: The Guilford Press, 1990. pp. 205–225.

McCornac, D.C., and Filante, R.W. The demand for distilled spirits: An empirical investigation. *J Stud Alcohol* 45:176–178, 1984.

McGuinness, T. The demand for beer, spirits and wine in the U.K., 1956–1979. In: Grant, M.; Plant, M.; and Williams, A., eds. *Economics and Alcohol: Consumption and Controls.* New York: Gardner, 1983. pp. 238–242.

Mulford, H.A., and Fitzgerald, J.L. The consequences of increasing off-premise wine outlets in Iowa. *Br J Addict* 83: 1271–1279, 1988.

Nelson, J.P. "Effects of Regulation on Alcoholic Beverage Consumption: Regression Diagnostics and Influential Data." Department of Economics, Pennsylvania State University, University Park, PA, 1988.

Neuman, C., and Rabow, J. Drinkers' use of physical availability of alcohol: Buying habits and consumption level. *Int J Addict* 20: 1663–1673, 1986.

Ornstein, S.I., and Hanssens, D.M. Alcohol control laws and the consumption of distilled spirits and beer. *J Consumer Res* 12:200–213, 1985.

Parker, R.N. Alcohol and theories of homicide. In: Adler, F., and Laufer, W.S., eds. *Advances in Criminological Theory.* Vol. 4. New Brunswick: Transaction Press, 1992. pp. 115–143.

Parker, D.A.; Wolz, M.W.; and Harford, T.C. The prevention of alcoholism: An empirical report on the effects of outlet availability. *Alcohol Clin Exp Res* 2:339–343, 1978.

Pindyck, R.S., and Rubinfeld, D.L. *Microeconomics.* New York: Macmillan Publishing Company, 1989.

Prendergast, M.L. A history of alcohol problem prevention efforts in the United States. In: Holder, H.D., ed. *Control Issues in Alcohol Abuse Prevention: Strategies for States and Communities.* Greenwich, CT: JAI Press, Inc., 1987. pp. 25–52.

Rabow, J.; Schwartz, C.; Stevens, S.; and Watts, R.K. Social psychological dimensions of alcohol availability: The relationship of perceived social obligations, price considerations, and energy expended to the frequency, amount, and type of alcoholic beverage consumed. *Int J Addict* 17:1259–1271, 1982.

Rabow, J., and Watts, R.K. Alcohol availability, alcoholic beverage sales and alcohol-related problems. *J Stud Alcohol* 43:767–801, 1982.

Room, R. "Alcohol Monopolies in the U.S.A.: Challenges and Opportunities." Paper presented at a meeting organized by the Swedish Systembolaget on the Role of Alcohol Monopolies, Vaxholm, Sweden, 1987.

Rush, B.R., and Gliksman, L. The distribution of consumption approach to the prevention of alcohol-related damage: An overview of relevant research and current issues. *Ad Alcohol Subst Abuse* 5:9–32, 1986.

Rush, B.R.; Gliksman, L.; and Brook, R. Alcohol availability, alcohol consumption and alcohol-related damage. I. The distribution of consumption model. *J Stud Alcohol* 47:1–10, 1986.

Saffer, H., and Grossman, M. Drinking age laws and highway mortality rates: Cause and effect. *Econ Inquiry* 25:403–417, 1987.

Saltz R.; Hennessy, M.; and Gruenewald, P.J. "Candidate Alcohol Problems and Implications for Measurement: General Alcohol Problems, Outcome Measures, Instrumentation and Surrogates." Paper prepared for the conference entitled Methodological Issues in Community Prevention Trials for Alcohol Problems, Berkeley, CA, 1989.

Schweitzer, S.O.; Intriligator, M.D.; and Salehi, H. Alcoholism: An econometric model of its causes, its effects and its control. In: Grant, M.; Plant, M.; and Williams, A., eds. *Economics and Alcohol: Consumption and Controls.* New York: Gardner, 1983. pp. 107–127.

Simon, J.L. The economic effects of state monopoly of packaged-liquor retailing. *J Political Econ* 74:188–194, 1966*a*.

Simon, J.L. The price elasticity of liquor in the U.S. and a simple method of determination. *Econometrica* 34:193–205, 1966*b*.

Single, E. The availability of alcohol: Prior research and future directions. *Aust Drug Alcohol Rev* 7:273–284, 1988.

Skog, O-J. The collectivity of drinking cultures: A theory of the distribution of alcohol consumption. *Br J Addict* 80:83–99, 1985.

Smart, R.G. The relationship of availability of alcoholic beverages to per capita consumption and alcoholism rates. *J Stud Alcohol* 38:891–896, 1977.

Smart, R.G. Availability and the prevention of alcohol-related problems. In: Harford, T.C.; Parker, D.A.; and Light, L., eds. *Normative Approaches to the Prevention of Alcohol Abuse and Alcoholism.* National Institute on Alcohol Abuse and Alcoholism Research Monograph 3. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1980. pp. 123–146.

Smart, R.G. The impact on consumption of selling wine in grocery stores. *Alcohol Alcohol* 21:233–236, 1986.

Smart, R.G., and Docherty, D. Effects of the introduction of on-premise drinking in alcohol-related accidents and impaired driving. *J Stud Alcohol* 37:683–686, 1976.

Smith, D.I. Effect on traffic accidents of introducing Sunday hotel sales in New South Wales, Australia. *Contemp Drug Prob* 14:279–295, 1987.

Smith, D.I. Effect on traffic accidents of introducing Sunday alcohol sales in Brisbane, Australia. *Int J Addict* 23:1091–1099, 1988.

Sprattler, K.M.; Parker, R.N.; and Gruenewald, P.J. Administrative license revocation statutes: Do they really deter drunk driving? *J Quant Criminol*, in review.

Stevens, J. *Applied Multivariate Statistics for the Social Sciences.* Hillsdale, NJ: Lawrence Erlbaum, 1986.

Swidler, S. A reexamination of liquor price and consumption differences between public and private ownership states: Comment. *South Econ J* July:259–268, 1986.

Tan, E.S.; Lemmens, P.H.H.M.; and Koning, A.J. Regularity in alcohol distributions: Implications for the collective nature of drinking behavior. *Br J Addict* 85:745–750, 1990.

Treno, A.J.; Nephew, T.; Ponicki, W.R.; and Gruenewald, P.J. Alcohol beverage price spectra from the state of California: Opportunities for substitution. *Alcohol Clin Exp Res*, in press.

Treno, A.J.; Parker, R.; and Holder, H.D. Understanding U.S. alcohol consumption with

social and economic factors: A multivariate time series analysis, 1950–1986. *J Stud Alcohol* 54(2):146–156, 1993.

Wagenaar, A.C., and Holder, H.D. A change from public to private sale of wine: Results from natural experiments in Iowa and West Virginia. *J Stud Alcohol* 52:162–173, 1991.

Walsh, B.M. The demand for alcohol in the UK: A comment. *J Ind Econ* 30:439–446, 1982.

Watts, R.K., and Rabow, J. Alcohol availability and alcohol-related problems in 213 California cities. *Alcohol Clin Exp Res* 7:47–58, 1983.

Williams, G.P. Off-licenses in supermarkets open a new door to liquor for many. *J Alcohol* 10:122–131, 1975.

Wilkinson, J.T. Reducing drunken driving: Which policies are most effective? *South Econ J* 54:322–334, 1987.

Wittman, F.D., and Hilton, M.E. Uses of planning and zoning ordinances to regulate alcohol outlets in California cities. In: Holder, H., ed. *Control Issues in Alcohol Abuse Prevention: Strategies for States and Communities.* Suppl. 1. Greenwich, CT: JAI Press, 1987. pp. 337–366.

Zardkoohi, A., and Sheer, A. Public versus private liquor retailing: An investigation into the behavior of the state governments. *South Econ J* July 1983-April 1984:1058–1076, 1984.

BIBLIOGRAPHY OF REVIEWED MATERIALS

Abbey, A.; Scott, R.O.; Oliansky, D.M.; Quinn, B.; and Andreski, P.M. Subjective, social, and physical availability. I. Their interrelationships. *Int J Addict* 25(8):889–910, 1990.

Abbey, A.; Scott, R.O.; Oliansky, D.M.; Quinn, B.; and Andreski, P.M. Subjective, social, and physical availability. II. Their simultaneous effects on alcohol consumption. *Int J Addict* 25(9):1011–1023, 1990.

Barrett, M.E.; Joe, G.W.; and Simpson, D.D. Availability of drugs and psychological proneness in opioid addiction. *Int J Addict* 25(10): 1211–1226, 1990.

Berg, L.L., and Colman, C.B. Patterns of consumption of beer and wine, retail availability, and DUI. *J Alcohol Drug Educ* 36:46–67, 1991.

Blose, J., and Holder, H.D. Liquor-by-the-drink and alcohol-related traffic crashes: A natural experiment using time series analysis. *J Stud Alcohol* 48:52–60, 1987.

Blose, J., and Holder, H. The public availability of distilled spirits: Structural and reported consumption changes associated with liquor-by-the-drink. *J Stud Alcohol* 48:371–379, 1987.

Bourgeois, J.C., and Barnes, J.G. Does advertising increase alcohol consumption? *J Advert Res* 19(4):19–29, 1979.

Brief, R.P. "The Relationship Between the Number of Off-Premise Licenses and the Consumption of Alcoholic Beverages: A Statistical Analysis." New York State Moreland Commission on Alcoholic Beverage Control Law, Study Paper #3. 1963.

Bruce, D. Changes in Scottish drinking habits and behaviour following the extension of permitted evening opening hours. *Health Bull* 38:133–137, 1980.

Bruun, K.; Edwards, G.; Lumio, M.; Mäkelä, K.; Pan, L.; Popham, R.E.; Room, R.; Schmidt, W.; Skog, O-J.; Sulkunen, P.; and Osterberg, E. *Alcohol Control Policies in Public Health Perspective.* Helsinki: Finnish Foundation for Alcohol Studies, 1975.

Bryant, C.W. Effects of sale of liquor by the drink in the State of Washington. *Q J Stud Alcohol* 15:320–324, 1954.

Cartwright, A.K.J.; Shaw, S.J.; and Spratley, T.A. The relationships between per capita consumption, drinking patterns and alcohol related problems in a population sample, 1965–1974. Part II: Implications for alcohol control policy. *Br J Addict* 73:247–258, 1978.

Colon, I. The influence of state monopoly of alcohol distribution and the frequency of package stores on single motor vehicle fatalities. *Am J Drug Alcohol Abuse* 9:325–331, 1982.

Colon, I., and Cutter, H.S.G. Relationship of beer consumption and state alcohol and motor vehicle policies to fatal accidents. *J Safety Res* 14:83–89, 1983.

Colon, I.; Cutter, H.S.G.; and Jones, W.C. Prediction of alcoholism from alcohol availability, alcohol consumption and demographic data. *J Stud Alcohol* 43(11):1199–1213, 1982.

Dewar, R., and Sommer, R. The consumption of alcohol in a Saskatchewan community before and after the opening of a new liquor outlet. Regina, Saskatchewan: Bureau on Alcoholism, 1962.

Dewar, R., and Sommer, R. Alcohol consumption in Prairieville. Effect of change in type of outlet on drinking in a small town. *Q J Stud Alcohol* 25:300–313, 1964.

Douglas, R.L.; Wagenaar, A.C.; and Barkey, P.M. The relationship of changing alcohol availability to acute and chronic social and health problems. In: Galanter, M., ed. *Recent Advances in Research and Treatment.* Currents in Alcoholism VII. New York: Grune and Stratton, 1980. pp. 401–428.

Dull, R.T., and Giacopassi, D.J. An assessment of the effects of alcohol ordinances on selected behaviors and conditions. *J Drug Issues* 4:511–521, 1986.

Dull, R.T., and Giacopassi, D.J. Dry, damp, and wet: Correlates and presumed consequences of local alcohol ordinances. *Am J Drug Alcohol Abuse* 14(4):499–514, 1988.

Ede, A., and Templar, D.I. The effect of a liquor store strike upon psychiatric admissions for alcoholism. *Can Psychiatr Assoc J* 21:501, 1976.

Entine, A.D. "The Relationship Between the Number of Sales Outlets and the Consumption of Alcoholic Beverages in New York and Other States." New York, NY: New York State Moreland Commission on the Alcoholic Beverage Control Law, Study Paper #2. 1963.

Fillmore, K.M., and Wittman, F.D. The effects of availability of alcohol on college student drinking: A trend study. *Contemp Drug Prob* 11:455–492, 1982.

Fitzgerald, J.L., and Mulford, H.A. Consequences of increasing alcohol availability: The Iowa experience revisited. *Br J Addict* 87:267–274, 1992.

Fosdick, R.B., and Scott, A.L. *Toward Liquor Control.* New York: Harper and Brothers, 1933.

Frankel, B.G., and Whitehead, P.C. *Drinking and Damage: Theoretical Advances and Applications for Prevention.* Monograph 14. New Brunswick, NJ: Rutgers Center for Alcohol Studies, 1981.

Giesbrecht, N., and MacDonald, S. A ban of fortified wine in Northwestern Ontario and its impact on the consumption level and drinking patterns. *Br J Addict* 76:281–288, 1981.

Giesbrecht, N.; Markle, G.; and MacDonald, S. The 1978–79 INCO workers' strike in the Sudbury basin and its impact on alcohol consumption and drinking patterns. *J Public Health Policy* 3:22–38, 1982.

Gliksman, L., and Rush, B.R. Alcohol availability, alcohol consumption and alcohol related damage. II. The role of sociodemographic factors. *J Stud Alcohol* 47:11–18, 1986.

Godfrey, C. Licensing and the demand for alcohol. *Appl Econ* 20:1541–1558, 1988.

Godfrey, C. Factors influencing the consumption of alcohol and tobacco: The use and abuse of economic models. *Br J Addict* 84:1123–1138, 1989.

Greenaway, J.R. The drink problem back on the political agenda? *Political Q* 61:80–92, 1990.

Grossman, M. "Health Economics of Prevention of Alcohol-Related Problems." Paper presented at the workshop on Health Economics of Prevention and Treatment of Alcohol-Related Problems, National Institute on Alcohol Abuse and Alcoholism, Washington, DC, 1988.

Grossman, M. Health benefits of increases in alcohol and cigarette taxes. *Br J Addict* 84:1193–1204, 1989.

Gruenewald, P.J. "A Choice Model of Alcohol Consumption." Paper presented at the Research Society on Alcoholism annual meeting, Beaver Creek, CO, 1989.

Gruenewald, P.J. Loss of control drinking among first offender drunk drivers. *Alcohol Clin Exp Res* 15(4):634–639, 1991.

Gruenewald, P.J.; Madden, P.; and Janes, K. Alcohol availability and the formal power and resources of state alcohol beverage control agencies. *Alcohol Clin Exp Res* 16:591-597, 1992.

Gruenewald, P.J., and Nephew, T. "Consumption and Death: The Dynamics of Alcohol Consumption Patterns." Paper presented at the annual meeting of the Research Society on Alcoholism, Marco Island, FL, 1991.

Gruenewald, P.J.; Ponicki, W.B.; and Holder, H.D. The relationship of outlet densities to alcohol consumption: A time series cross-sectional analysis. *Alcohol Clin Exp Res* 17(1): 38–47, 1993.

Hammer, T., and Vaglu, P. The increase in alcohol consumption among women: A phenomenon related to accessibility or stress? A general population study. *Br J Addict* 84:767–775, 1989.

Hammond, R.L. Do increased alcohol outlets influence alcoholism rates? *Rep Alcohol* 33:12–19, 1975.

Harford, T.C.; Parker, D.A.; Pautler, C.; and Woltz, M. Relationship between the number of on-premise outlets and alcoholism. *J Stud Alcohol* 40:1053–1057, 1979.

Hauge, R. The effects of changes in availability of alcoholic beverages. In: Laurence, M.D.; Snortum, J.R.; and Zimring, F.E., eds. *Social Control of the Drinking Driver.* Chicago: The University of Chicago Press, 1988. pp. 169–187.

Hoadley, J.F.; Fuchs, B.C.; and Holder, H.D. The effect of alcohol beverage restrictions on consumption: A 25-year longitudinal analysis. *Am J Drug Alcohol Abuse* 10(3):375–401, 1984.

Holder, H.D., and Blose, J. Impact of changes in distilled spirits availability on apparent consumption: A time series analysis of liquor-by-the-drink. *Br J Addict* 82:623–631, 1987.

Holder, H.D., and Wagenaar, A.C. Effects of the elimination of a state monopoly on distilled spirits' retail sales: A time series analysis of Iowa. *Br J Addict* 85:1615–1625, 1991.

Horverak, O. The 1978 strike at the Norwegian Wine and Spirits Monopoly. *Br J Addict* 78:51–66, 1983.

Janes, K., and Gruenewald, P.J. The role of formal law in alcohol control systems: A comparison among states. *Am J Drug Alcohol Abuse* 17:199–214, 1991.

Kendell, R.E.; de Roumanie, M.; and Ritson, E.B. Effect of economic changes on Scottish drinking habits, 1978–82. *Br J Addict* 78:365–379, 1983.

Kiholm Smith, J. An analysis of state regulations governing liquor store licenses. *J Law Econ* 25:301–319, 1982.

Kuusi, P. *Alcohol Sales Experiment in Rural Finland.* Vol 3. Helsinki: Finnish Foundation of Alcohol Studies, 1957.

Lacey, J.H.; Stewart, J.R.; and Carroll, C.L. An examination of the effect of liquor-by-the-drink on traffic safety in North Carolina. In: Goldberg, L., ed. *Alcohol, Drugs and Traffic Safety.* Stockholm: Almqvist and Wiksell, 1981. pp. 1382–1388.

Lenke, L. Alcohol and crimes of violence: A causal analysis. *Contemp Drug Prob* Fall: 355–365, 1982.

Levin, H.J. Economic and regulatory aspects of liquor licensing. *Univ Pennsylvania Law Rev* 112(6):785–833, 1964.

Lind, B., and Herbert, D.C. The effect of "Sunday trading" on traffic crashes. Traffic

Accident Research Unit, Traffic Authority of New South Wales, SR 82/112, 1982.

Loeb, B.F., Jr. Relationship of state law to per capita drinking. In: Ewing, J.A., and Rouse, B.A., eds. *Drinking-Alcohol in American Society.* Chicago: Nelson-Hall, 1978. pp. 219–239.

MacDonald, S. The impact of increased availability of wine in grocery stores on consumption: Four case histories. *Br J Addict* 81:381–387, 1986.

MacDonald, S., and Whitehead, P.C. Availability of outlets and consumption of alcoholic beverages. *J Drug Issues* 13:477–486, 1983.

Magruder, K.M. The association of alcoholism mortality with legal availability of alcoholic beverages. *J Alcohol Drug Educ* 21:27–37, 1976.

Mäkelä, K. *Types of Alcohol Restrictions, Types of Drinkers and Types of Alcohol Damages: The Case of the Personnel Strike in the Stores of the Finnish Alcohol Monopoly.* Finnish Social Research Institute of Alcohol Studies, Report No. 77, 1974.

Mäkelä, K. Differential effects of restricting the supply of alcohol: Studies of a strike in Finnish liquor stores. *J Drug Issues* 10:131–144, 1980.

Malmquist, S. *A Statistical Analysis of the Demand for Liquor in Sweden; A Study of the Demand for a Rationed Commodity.* Uppsala: Appelbergs Boktryckeri, 1948.

Mann, R.E., and Anglin, L. Alcohol availability, per capita consumption, and the alcohol-crash problem. In: Wilson, R.J., and Mann, R.E., eds. *Drinking and Driving: Advances in Research and Prevention.* New York: The Guilford Press, 1990.

McCornac, D.C. and Filante, R.W. The demand for distilled spirits: An empirical investigation. *J Stud Alcohol* 45: 176–178, 1984.

McGuinness, T. An econometric analysis of total demand for alcoholic beverages in the U.K. *J Ind Econ* 29:85–109, 1980.

McGuinness, T. The demand for beer, spirits and wine in the U.K., 1956–79. In: Grant, M.; Plant, M.; and Williams, A., eds. *Economics and Alcohol: Consumption and Controls.* New York: Gardner, 1983. pp. 238–242.

Medicine in the Public Interest, Inc. *The Effects of Alcoholic-Beverage-Control Laws.* Washington, DC: Medicine in the Public Interest, Inc., 1979.

Mulford, H.A., and Fitzgerald, J.L. The consequences of increasing off-premise wine outlets in Iowa. *Br J Addict* 83:1271–1279, 1988.

Nelson, J.P. "Effects of Regulation on Alcoholic Beverage Consumption: Regression Diagnostics and Influential Data." Department of Economics, Pennsylvania State University, University Park, PA, 1988.

Neuman, C., and Rabow, J. Drinkers' use of physical availability of alcohol: Buying habits and consumption level. *Int J Addict* 20:1663–1673, 1986.

Ogborne, A.C., and Smart, R.G. Will restrictions on alcohol advertising reduce alcohol consumption? *Br J Addict* 75:293–296, 1980.

Olsson, O., and Wikstrom, P-O.H. Effects of the experimental Saturday closing of liquor retail stores in Sweden. *Contemp Drug Prob* 11:325–353, 1982.

Ornstein, S.I., and Hanssens, D.M. Alcohol control laws and the consumption of distilled spirits and beer. *J Consumer Res* 12: 200–213, 1985.

Osterberg, E. "The 1972 and 1985 Finnish Alcohol Retail Outlet Strikes." Presented at Annual Meeting of Kettil Bruun Society for Social and Epidemiological Research on Alcohol, Berkeley, CA, 1988.

Osterberg, E. Would a more liberal control policy increase alcohol consumption? *Contemp Drug Prob* Winter:545–573, 1990.

Parker, D.A., and Wolz, M. Alcohol problems and the availability of alcohol. *Alcohol Clin Exp Res* 3:309–312, 1979.

Parker, D.A.; Wolz, M.W.; and Harford, T.C. The prevention of alcoholism: An empirical

report on the effects of outlet availability. *Alcohol Clin Exp Res* 2:339–343, 1978.

Popham, R.E. The urban tavern: Some preliminary remarks. *Addictions* 9(2):16–28, 1962.

Popham, R.E.; Schmidt, W.; and de Lint, J. The effect of legal restraints on drinking. In: Kissin, B., and Begleiter, J., eds. *The Biology of Alcoholism, Social Aspects of Alcohol.* Vol. 4. New York: Plenum Press, 1976. pp. 579–625.

Popham, R.E.; Schmidt, W.; and de Lint, J. Government control measures to prevent hazardous drinking. In: Ewing, J.A., and Rouse, B.A., eds. *Drinking-Alcohol in American Society.* Chicago: Nelson-Hall, 1978. pp. 239–266.

Rabow, J.; Schwartz, C.; Stevens, S.; and Watts, R.K. Social psychological dimensions of alcohol availability: The relationship of perceived social obligations, price considerations, and energy expended to the frequency, amount, and type of alcoholic beverage consumed. *Int J Addict* 17(8):1259–1271, 1982.

Rabow, J., and Watts, R.K. Alcohol availability, alcoholic beverage sales and alcohol-related problems. *J Stud Alcohol* 43: 767–801, 1982.

Raymond, A. Ten o'clock closing—the effect of the change in hotel bar closing time on road accidents in the metropolitan area of Victoria. *Australia Road Res* 3(10):3–17, 1969.

Rush, B.R., and Gliksman, L. The distribution of consumption approach to the prevention of alcohol-related damage: An overview of relevant research and current issues. *Adv Alcohol Subst Abuse* 5:9–32, 1986.

Rush, B.R.; Gliksman, L.; and Brook, R. Alcohol availability, alcohol consumption and alcohol-related damage. I. The distribution of consumption model. *J Stud Alcohol* 47:1–10, 1986.

Rush, B.R.; Steinberg, M.; and Brook, R. The relationships among alcohol availability, alcohol consumption and alcohol-related damage in the province of Ontario and the state of Michigan 1955–1982. *Adv Alcohol Subst Abuse* 5(4):33–45, 1986.

Saila, S.L. "A Trial Closure of Alko Retail Outlets on Saturdays and Its Effects on Alcohol Consumption and Disturbances Caused by Intoxication." Paper presented at the 24th International Institute on the Prevention and Treatment of Alcoholism. Lausanne: International Council on Alcohol and Addictions, 1978.

Simon, J.L. The economic effects of state monopoly of packaged-liquor retailing. *J Political Econ* 74:188–194, 1966.

Single, E. The availability of alcohol: Prior research and future directions. *Aust Drug Alcohol Rev* 7:273–284, 1988.

Smart, R.G. The effect of licensing restrictions during 1914–18 on drunkenness and liver cirrhosis deaths in Britain. *Br J Addict* 69:109–121, 1974.

Smart, R.G. The effects of two liquor store strikes on drunkenness, impaired driving and accidents. *J Stud Alcohol* 38: 1785–1789, 1977.

Smart, R.G. The relationship of availability of alcoholic beverages to per capita consumption and alcoholism rates. *J Stud Alcohol* 38:891–896, 1977.

Smart, R.G. Availability and the prevention of alcohol-related problems. In: Harford, T.C.; Parker, D.A.; and Light, L., eds. *Normative Approaches to the Prevention of Alcohol Abuse and Alcoholism.* National Institute on Alcohol Abuse and Alcoholism Research Monograph 2. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1980. pp. 123–146.

Smart, R.G. The impact on consumption of selling wine in grocery stores. *Alcohol* 21:233–236, 1986.

Smart, R.G. Socio-economic, lifestyle, and availability factors in the stabilization of alcohol consumption in Canada. *Can J Public Health* 78:176–180, 1987.

Smart, R.G., and Adlaf, E.M. Banning happy hours: The impact of drinking and impaired-driving charges in Ontario, Canada. *J Stud Alcohol* 47:256–258, 1986.

Smart, R.G., and Docherty, D. Effects of the introduction of on-premise drinking on alcohol-related accidents and impaired driving. *J Stud Alcohol* 37:683–686, 1976.

Smith, D.I. Impact on traffic safety of the introduction of Sunday alcohol sales in Perth, Western Australia. *J Stud Alcohol* 39:1302–1304, 1978.

Smith, D.I. Comparison of patrons of hotels with early opening and standard hours. *Int J Addict* 2:155–163, 1986.

Smith, D.I. Effect on traffic accidents of introducing Sunday hotel sales in New South Wales, Australia. *Contemp Drug Prob* 14:279–295, 1987.

Smith, D.I. Effect on traffic accidents of introducing flexible hotel trading hours in Tasmania, Australia. *Br J Addict* 83: 219–222, 1988.

Smith, D.I. Effect on traffic accidents of introducing Sunday alcohol sales in Brisbane, Australia. *Int J Addict* 23: 1091–1099, 1988.

Smith, D.I. Effectiveness of restrictions on availability as a means of preventing alcohol-related problems. *Contemp Drug Prob* Winter:627–684, 1988.

Smith, D.I. Effect on casualty traffic accidents of the introduction of 10 p.m. Monday to Saturday hotel closing in Victoria. *Aust Drug Alcohol Rev* 7:163–166, 1988.

Spellman, W.E., and Jorgenson, M.R. Liquor control and consumption. *J Stud Alcohol* 44:194–197, 1983.

Wagenaar, A.C., and Holder, H.D. A change from public to private sale of wine: Results from natural experiments in Iowa and West Virginia. *J Stud Alcohol* 52:162–173, 1991.

Walsh, B.M. The demand for alcohol in the UK: A comment. *J Indust Econ* 30:439–446, 1982.

Watts, R.K.; Phil, C.; and Rabow, J. The role of tourism in measures of alcohol consumption, alcohol availability and alcoholism. *J Stud Alcohol* 42:797–801, 1981.

Watts, R.K., and Rabow, J. Alcohol availability and alcohol-related problems in 213 California cities. *Alcohol Clin Exp Res* 7:47–58, 1983.

Whitehead, P.C. Effects of liberalizing alcohol control measures. *Addict Behav* 1:197–203, 1976.

Williams, G.P. Off-licenses in supermarkets open a new door to liquor for many. *J Alcohol* 10:122–131, 1975.

Wilkinson, J.T. Reducing drunken driving: Which policies are most effective? *South Econ J* 54:322–334, 1987.

Wittman, F.D., and Hilton, M.E. Uses of planning and zoning ordinances to regulate alcohol outlets in California cities. In: Holder, H., ed. *Control Issues in Alcohol Abuse Prevention: Strategies for States and Communities.* Suppl. 1. Greenwich, CT: JAI Press, 1987. pp. 337–366.

Zardkoohi, A., and Sheer, A. Public versus private liquor retailing: An investigation into the behavior of the state governments. *South Econ J* July 1983-April 1984:1058–1076, 1984.

4

The Economic Analysis of Addictive Behavior

Michael Grossman, Ph.D.

INTRODUCTION

My tasks in this paper are to review and critically evaluate economic models of addictive behaviors, particularly harmfully addictive behaviors. Part of my discussion includes topics for continuing research attention. Examples of the behaviors at issue are cigarette smoking; gambling; consumption of such illegal drugs as heroin, marijuana, and cocaine; and excessive alcohol use. Given the theme of this volume, I focus on current and future research in the last area, although I use recently completed studies of cigarette smoking and gambling at horse-racing tracks to highlight the empirical relevance of economic analyses of addiction, especially models that treat addiction as rational behavior.

At the outset, I want to provide some information concerning my research background in the areas that I treat in this paper. I have been engaged in research on the demand for cigarettes and alcohol for more than a decade (Lewit et al. 1981; Grossman et al. 1987; Saffer and Grossman 1987*a*, 1987*b*; Coate and Grossman 1988; Grossman 1989; Becker et al. 1991*a*, 1991*b*; Chaloupka et al. in press; Grossman et al. in progress). Although cigarette smoking and excessive alcohol use are obvious examples of harmfully addictive behaviors, my initial interest in these behaviors emanated from a public health policy perspective rather than from a desire to contribute to the theory of addiction. Given the mounting evidence that lifestyles play a more important role in health outcomes than medical care and the obvious detrimental health effects of the behaviors at issue, I reasoned that policymakers would be interested in empirical evidence of behaviors with regard to their sensitivity to price—a policy-manipulable variable in light of excise taxes—and other Government regulations such as the minimum legal drinking age for the purchase and consumption of alcoholic beverages.

I was particularly interested in price as a determinant of alcohol and cigarette consumption by teenagers and young adults. The desire to estimate demand functions for a particular segment of the population required the use of micro data, where one can capitalize on the substantial cross-sectional variation in alcohol and cigarette prices due primarily to differences in State excise tax rates on these goods. It also led my colleagues at the

National Bureau of Economic Research and me to publish the first studies dealing with the effects of cigarette prices on smoking and with the effects of alcohol prices on alcohol use in micro data.

My hypothesis that price might be an important determinant of teenage cigarette smoking and alcohol use was treated with a good deal of skepticism by Initial Review Groups at the National Institute of Child Health and Human Development and the National Institute on Alcohol Abuse and Alcoholism (NIAAA) in 1979. It also was viewed with skepticism by a number of economists who accepted the conventional wisdom that the demand for addictive substances is unlikely to be responsive to price. My hypothesis on cigarette smoking received more sympathy from National Science Foundation reviewers in 1980, and a proposal dealing with the effects of Government regulation (including price regulation) on teenage alcohol use and motor vehicle accident mortality received similar treatment from NIAAA reviewers in 1982. In my view, my success at NIAAA was due in part to Philip Cook's seminal contributions to the literature on the economics of alcohol prevention problems (Cook and Tauchen 1982, 1984), which began to receive widespread attention around the time that my NIAAA proposal was evaluated.

In 1986, Gary Becker and Kevin Murphy invited me to join them in a research project that would test empirically their theory of rational addiction. To date, this collaboration has resulted in a paper pertaining to cigarette addiction in a time series of State cross sections (Becker et al. 1991*a*), a second paper containing some theoretical extensions of the rational addiction model (Becker et al. 1991*b*), and an NIAAA-funded project dealing with alcohol, addiction, and price (Grossman et al. in progress). In addition, I have supervised a Ph.D. dissertation by Frank Chaloupka (1988) that applies the rational addiction model to cigarette smoking in micro data[1] and a Ph.D. dissertation by Pamela Mobilia (1990) that applies the same model to the demand for gambling at horse-racing tracks in a time series of track cross-sections. Finally, I have developed a research proposal to estimate demand functions for cocaine, marijuana, and heroin using both conventional models of consumer behavior and models of addictive behavior. So far my attempts to fund this project have not been successful, but given my "rational addiction" to writing proposals, that situation may change.

I provided some information on my background in alcohol and addiction research to enable readers of this paper to put the views expressed in it in the proper perspective. First and foremost, my evaluation of alternative economic theories of addiction is based on their ability to contribute to an understanding of behavior, particularly price sensitivity, at the empirical level. I originally approached the demand for alcohol and cigarettes in the context of a conventional model of consumer behavior, and I applaud current

[1] *The main results of Chaloupka's dissertation are published in Chaloupka (1991).*

efforts by Manning et al. (1991), Kenkel (in press), and Phelps and Leung (in progress) to refine estimates of demand functions for alcoholic beverages in micro data regardless of the particular theoretical perspective taken in each study. Indeed, in my view the main contribution of Becker and Murphy's (1988) model of rational addiction is to suggest that it is a mistake to assume that addictive goods are not sensitive to price. Even if one does not accept all the aspects of their model, one can examine this proposition in the context of the standard theory of consumer behavior. Second, I am clearly not an unbiased evaluator of the rational addiction model relative to other economic models of addictive behavior. In my opinion the alternative models are characterized by needless complexity and/or lack of clear empirical implications. But I am not the appropriate person to provide a balanced discussion of this issue since I am involved with most of the empirical tests of the rational addiction model. Moreover, empirical research using a rational addiction framework is very new, while similar research using a myopic addiction framework is relatively old. I want to focus on the new contributions in this area. For these reasons, the reader will want to consult the references that I list to evaluate fully competing economic analyses of addiction.

The plan of the remainder of this paper is as follows. In the first section, I indicate how addiction is defined in economic approaches to this behavior and also define some key concepts used in these approaches. In the next section, I briefly review theoretical and empirical economic contributions in this area that either assume that consumers are myopic or that they have stable but inconsistent short-run and long-run preferences. Then, I outline theories of rational addiction with a focus on Becker and Murphy's (1988) contribution and extensions developed by Becker et al. (1991*b*). In the next section, I describe empirical applications of rational addiction to the demand for cigarettes, gambling at horse-racing tracks, and leisure time. Then, I treat empirical evidence of the relevance of the rational addiction model in studying the demand for excessive alcohol consumption. Included here are preliminary results from my research with Becker, Chaloupka, and Murphy (Grossman et al. in progress) in which we estimate rational addiction demand functions for excessive alcohol consumption using liver cirrhosis mortality as the outcome measure. Finally, I consider topics for continuing research attention.

DEFINITIONS AND BASIC CONCEPTS

In economic analyses of addictive behavior, the consumption of a certain good is termed to be an addiction or a habit if an increase in current consumption of the good leads to an increase in future consumption. Of course, this means that current consumption is positively related to past consumption. A harmful addiction is one in which current consumption has detrimental effects in the future, such as reductions in health and therefore in utility caused by cigarette smoking and excessive

alcohol use. Similarly, a beneficial addiction is one in which current consumption has favorable effects in the future, such as increases in future utility from attending and acquiring additional information about operas today. Other examples of beneficial addictions are jogging or playing tennis on a regular basis.[2] Note that I follow the literature in using the terms addiction and habit as synonyms.[3]

Consumers are myopic if they ignore the effects of current consumption on future utility when they determine the optimal or utility-maximizing quantity of an addictive good in the present period. On the other hand they are rational or farsighted if they take account of future effects of current consumption when they determine the optimal quantity of an addictive good in the present period.

Both myopic and rational models of addiction stress that choices today depend on choices made in the past. The usual assumption in most treatments of behavior over time is that this is not the case. This assumption is built into models of optimal consumption over time or over the lifecycle by assuming that consumers maximize a utility function (V) that is additively separable over time. This utility function has the form

$$V = \sum_{t=1}^{\infty} \beta^{t-1} U(C_t) \qquad (1)$$

Here, C_t is consumption of a composite good at time t, $U(C_t)$ is the current period concave utility function at time t, and β is the time discount factor [$\beta = 1/(1 + r)$, where r is the rate of time preference for the present]. Thus, lifetime utility (V) is the discounted sum of utility in each period.

If equation 1 is maximized subject to a lifetime wealth constraint and if the rate of interest is equal to the rate of time preference for the present, the optimal quantity of C_t can be expressed as a function of the marginal utility of wealth (λ) and the price of C_t (P_t), as follows:

$$C_t = C_t(\lambda, P_t) \qquad (2)$$

Equation 2 specifies a demand function that holds the marginal utility of wealth constant. This is the standard demand function that is employed in lifecycle utility maximization models. Its properties and the specification of lifetime utility as equation 1 are discussed in detail by Hicks (1965), Heckman (1974), Ghez and Becker (1975), Deaton and Muellbauer (1980), MaCurdy (1981), and Phlips (1983). For our purposes it is important

[2] *Rimm et al. (1991) find that moderate alcohol consumption lowers the risk of coronary heart disease in men. Even if current and past moderate alcohol consumption are positively related, it does not follow that this type of consumption is another example of beneficial addiction. The reason is that moderate alcohol consumption may lead to such negative outcomes as strokes and motor vehicle accidents.*

[3] *An exception is Becker (1991). He defines an addiction as "a strong habit. Technically, a habit becomes an addiction when the effects of past consumption on present consumption are strong enough to be destabilizing" (p. 4). This produces an unstable steady state that can characterize certain equilibrium positions in the Becker-Murphy (1988) rational addiction model. I do not draw this distinction between a habit and an addiction in the text, but I do comment on unstable steady states in the third and fourth sections of this paper.*

to note that equations 1 and 2 rule out addiction because an increase in consumption in period t affects utility in period t alone.

To allow for the possibility that at least some goods are addictive, the utility function at time t is rewritten as

$$U = U(Y_t, C_t, S_t) \qquad (3)$$

where Y_t is consumption of the nonaddictive good, C_t is consumption of the addictive good, and S_t is the stock of the addictive good. The simplest specification of a process determining the stock at time t is

$$S_t = (1 - \delta)S_{t-1} + C_{t-1} \qquad (4)$$

where δ is the rate of depreciation on the stock. An increase in the stock lowers utility if the addiction is harmful ($\partial U/\partial S_t = U_S < 0$), while an increase in the stock raises utility if the addiction is beneficial ($U_S > 0$). Regardless of the nature of the addiction, an increase in the stock must raise the marginal utility of C_t in order for an increase in past consumption of C to increase current consumption. Becker and Murphy (1988) show that this is a sufficient condition for addiction in the case of myopic consumers, but it is not a sufficient condition for addiction in the case of rational consumers. I return to this point in the third section of this paper.

Experimental studies of harmful addiction by psychologists usually have identified reinforcement, tolerance, and withdrawal (Donegan et al. 1983; Peele 1985). Reinforcement means that greater current consumption of a good raises its future consumption. Tolerance means that given levels of consumption are less satisfying when past consumption has been greater. Withdrawal refers to the negative physical reaction and other losses in satisfaction as consumption is terminated. These insights are incorporated into the economic approach to addiction just described. In particular, in the case of harmful addiction (the relevant case for excessive alcohol consumption, cigarettes, and illegal drugs), an increase in the stock *lowers* utility (tolerance) but *raises* the marginal utility of current consumption of C (reinforcement). Withdrawal also is present because a reduction in C lowers current utility. Reinforcement and withdrawal, but not tolerance, also are present in the case of beneficial addiction.

MYOPIC MODELS OF ADDICTION

Economists' interest in addictive behavior dates to Alfred Marshall. Writing in 1920, he noted, "Whether a commodity conforms to the law of diminishing or increasing return, the increase in consumption arising from a fall in price is gradual; and, further, habits which have once grown up around the use of a commodity while its price is low are not quickly abandoned when its price rises again" (Marshall 1920, p. 807). Most economists who have studied this behavior since Marshall have assumed myopia or imperfectly rational behavior. I classify two types of models under the general rubric of myopia. In one, consumers ignore the effects of current consumption on future utility when they determine the

optimal quantity of an addictive good in the present period. In this class of models, past consumption affects current consumption through an accumulated stock of habits. Economists who use this type of model frequently view it as one in which tastes are endogenous since the stock measure in the current period utility function given by equation 3 depends on past decisions (for example, Houthakker and Taylor 1970; Pollak 1970, 1976; von Weizsäcker 1971; El-Safty 1976; Hammond 1976). In the second class of models, consumers have stable but inconsistent short-run and long-run preferences (for example, Schelling 1978, 1984*a*, 1984*b*; Elster 1979; Winston 1980; Thaler and Shefrin 1981). At any one point in time the individual is both a "far-sighted planner and a myopic doer" (Thaler and Shefrin 1981, p. 392), with the planner and the doer in conflict.

Excellent reviews of the literature on myopic models of addiction are contained in Deaton and Muellbauer (1980), Phlips (1983), Iannaccone (1984), and Chaloupka (1988). I will not outline these models here. Instead, I want to focus on their applications. My discussion pertains to the first class of models since those in which consumers have "split personalities" have not to my knowledge been used in empirical studies of addiction by economists.

Most empirical applications of myopic addiction are based on the pioneering work by Houthakker and Taylor (1970). They argue that the stock of a commodity should have a positive impact on its current consumption in the presence of habit formation if the commodity is nondurable. If it is durable, a negative stock effect is possible due to inventory adjustment.[4] Houthakker and Taylor fit dynamic time-series demand functions for the United States and several countries in Western Europe using nonadditive and additive specifications. In the latter the demand functions must satisfy the budget constraint in the sense that projections from it must sum to total expenditures.

Houthakker and Taylor find a considerable amount of evidence in support of habit formation. Structural stock coefficients, which are retrieved from coefficients of lagged consumption and other lagged variables, are positive for almost all nondurable goods. Among the figures for 81 goods in the United States for the period 1929–64, the stock coefficients of alcoholic beverages and cigarettes are 1.1 and 0.9, respectively. The alcohol coefficient ranks 8th and the cigarette coefficient ranks 10th among the 81 coefficients obtained. Using a somewhat refined version of the Houthakker-Taylor model, Phlips (1972, 1978) finds that habit formation dominates U.S. time-series data. He does not treat alcohol and cigarettes as separate consumption items.

Several researchers have applied variants of the Houthakker-Taylor model to the demand for alcohol and cigarettes in the United Kingdom and Australia. Kennedy et al. (1973) and Thom (1984) find evidence of habit formation in their estimates of the demand for beer, wine,

[4] *In the case of a durable good, the stock effect is negative given inventory adjustment but no addiction.*

and distilled spirits in Ireland. Duffy (1980), however, finds no support of the Houthakker-Taylor habit formation hypothesis in beer, wine, and spirits demand equations for the United Kingdom as a whole. Tsolakis and collaborators (1983) report that the model is compatible with Australian wine consumption but not beer consumption. Jones (1989) reports that a habit model outperforms a nonhabit model for beer in the United Kingdom but not for spirits, wine, and cigarettes.

Studies by Grabowski (1976), Johnson and Oksanen (1977), and Baltagi and Levin (1986) contain lagged alcohol and cigarette consumption effects in the context of demand specifications that are somewhat more flexible and less demanding on the data than the specifications required to fit the Houthakker-Taylor model. Grabowski (1976) obtains demand functions for 15 goods, including alcoholic beverages but excluding cigarettes, using U.S. data for the period 1956–72. In his many alternative specifications the lagged alcohol consumption coefficient always is positive and statistically significant. The coefficient ranges between 0.5 and 0.7 and typically is the second or third largest lagged effect reported by Grabowski. Johnson and Oksanen (1977) fit demand functions for beer, wine, and spirits in Canada. Their data consist of a time series of province cross sections for the period 1956–70. Lagged consumption coefficients are positive for each beverage and significant for spirits and wine. The wine coefficient is approximately 0.5, the spirits coefficient is 0.3, and the beer coefficient is 0.1. Using a U.S. time series of State cross-sections for the period 1964–80 to fit cigarette demand functions, Baltagi and Levin (1986) report a lagged consumption coefficient of 0.9.

Since all of the studies just described assume myopia, they all treat lagged consumption or the stock as exogenous or predetermined. Even if this is appropriate, the coefficient of the lagged dependent variable is biased and inconsistent in the presence of serial correlation. While some of the studies attempt to correct for serial correlation, biases still may be present if the process that generates this phenomenon is misspecified. The alcohol demand studies by Thom (1984) and Jones (1989) combine Deaton and Muellbauer's (1980) Almost Ideal Demand System and the Houthakker-Taylor model, which results in very complex empirical specifications. Moreover, the functional forms and estimation methods used by Houthakker and Taylor (1970) and by Phlips (1972, 1978) are more complicated than the procedures used by Becker et al. (1991*a*) and described in this paper in the section entitled Empirical Applications of Rational Addiction.

A final criticism of myopic studies of addiction pertains to the role of advertising in the demand function. If advertising has lagged effects on consumption and if these effects can be modeled as a Koyck process, then the coefficient of lagged consumption reflects both the effect of habit formation and the effect of advertising.[5] I

[5] *I owe this point to Henry Saffer.*

will not discuss this issue in detail because it is the subject of Henry Saffer's paper in this volume. But I will make a few additional remarks on advertising in the context of rational addiction.

So far I have discussed empirical studies of myopic addiction that focus on the impact of past consumption on current consumption. Myopic studies of the demand for cigarettes in the United States by Young (1983) and in Finland by Pekurinen (1989) have a different emphasis. These studies attempt to implement Scitovsky's (1976) notion that consumers should respond more to price decreases than to price increases in the case of addictive goods. Although they present empirical evidence in support of this proposition, no one has developed a formal model that contains a proof of it. Moreover, both studies neglect any effects of past consumption on current consumption. In addition, the phenomena that they measure may be due to past and future price effects which play an important role in the demand for addictive goods (see Becker et al. 1991*a* and the third and fourth sections of this paper). In the context of Becker and Murphy's (1988) treatment of unstable steady states, asymmetric price responses in aggregate data are possible, although the response to a price reduction is not necessarily larger than the response to a price increase.

RATIONAL MODELS OF ADDICTION

In rational models of addiction, consumers take account of future effects of current consumption when they determine the optimal quantity of an addictive good in the present period. Important theoretical contributions in this area have been made by Ryder and Heal (1973), Boyer (1978, 1983), and Kydland and Prescott (1982). These authors focus on the optimal path of consumption over time when the utility function is not additively separable over time. That is, the lifetime utility function is given by the discounted sum of current period utility functions given by equation 3 rather than by equation 1. Excellent summaries of these models are contained in Iannaccone (1984) and Chaloupka (1988). They are used by Becker and Murphy (1988) as the point of departure for a specific model of rational addiction that among other things contains the first explicit derivation of long- and short-run demand function for addictive goods in the case of farsighted consumers. Contrary to conventional wisdom, Becker and Murphy stress that the demand for addictive goods is likely to be quite responsive to price.

Becker and Murphy assume that consumers maximize a lifetime utility function of the form

$$V = \sum_{t=1}^{\infty} \beta^{t-1} U(Y_t, C_t, S_t) \qquad (5)$$

where all variables are defined in the first section of this paper. Note that the current period utility function, U, is well behaved in the sense that an indifference curve between C_t, the addictive good, and Y_t, the nonaddictive good, is convex to the origin. Note also that the lifetime utility function is separable over time in Y_t, C_t

and the stock of the addictive good, S_t; however, it is not separable in C_t and Y_t alone.

Although the utility function is separable over time, the quantity demanded of the addictive good is inversely related not only to the current price of the good but also to its past and future price. That is, cross-price effects are negative, or the quantities of an addictive good consumed in different periods are complements. (These predictions pertain to a demand function that holds the marginal utility of wealth constant. As indicated in the first section of this paper, this is the standard demand function employed in lifecycle utility maximization models.) Past prices are relevant because they affect past consumption of the addictive good, and an increase in past consumption due to a reduction in past prices raises current consumption. The future price is relevant because its reduction raises future consumption which lowers the "shadow price" of current consumption. Stated differently, the demand function of a rational consumer exhibits the property of symmetry: Increases in past or future consumption for whatever reason cause current consumption to rise. It also follows that the long-run response of consumption to a permanent price change exceeds the short-run response in absolute value in the case of an addictive good. (These responses are defined below.) Since this property does not hold for a nonaddictive good, the price elasticity of demand is expected to be larger for the former than for the latter.

When the utility function is quadratic and the rate of depreciation on the addictive stock is one,[6] Becker and coinvestigators (1991*a*) show that the Becker-Murphy model generates a structural demand function for consumption of C of the form

$$C_t = \alpha C_{t-1} + \beta\alpha C_{t+1} + \alpha_1 P_t + \alpha_2 e_t + \alpha_3 e_{t+1} \qquad (6)$$

Here, P_t is the price of C_t, $\beta < 1$ is the time discount factor, e_t and e_{t+1} reflect the impact of unmeasured exogenous lifecycle variables on utility, and the intercept is suppressed. Since α is positive and α_1 is negative, current consumption is positively related to past and future consumption (C_{t-1} and C_{t+1}, respectively) and negatively related to current price. In particular, α measures the effect of an increase in past consumption on the marginal utility of current consumption. By symmetry, it also measures the effect of an increase in future consumption on the marginal impact of current consumption on next period's utility. The larger the value of α, the greater is the degree of reinforcement or addiction.

Equation 6 is the basis of the empirical implementation of the model in most of the studies described in the fourth section of this paper. Note that ordinary least-squares estimation of the equation would lead to biased estimates of the parameters of interest. The unobserved variables that affect utility in each period are likely to be serially correlated. Even if

[6] *If* $\delta = 1$, $S_t = C_{t-1}$.

these variables are uncorrelated, C_{t-1} depends on e_t and C_{t+1} depends on e_{t+1} through the optimizing behavior. These relationships imply that an ordinary least-squares estimation of the equation might incorrectly imply that past and future consumption affect current consumption, even when the true value of α is zero. Stated differently, past and future consumption are endogenous variables and must be treated as such in estimating the model.

Fortunately, the specification in equation 6 suggests a way to solve the endogeneity problem. The equation implies that current consumption is independent of past and future prices when past and future consumption are held constant; any effect of past or future prices on current consumption must derive from their effects on past or future consumption. Provided that the unobservables are uncorrelated with prices in these periods, past and future prices are logical instruments for past and future consumption, since past prices directly affect past consumption and future prices directly affect future consumption. Therefore, the empirical strategy amounts to estimating equation 6 by two-stage least squares, with past and future prices serving as instrumental variables for past and future consumption.

The statistical significance of the coefficient of future consumption provides a direct test of a rational model of addiction against an alternative model in which consumers are myopic. In the latter model they fail to consider the impact of current consumption on future utility and future consumption. That is, the myopic version of equation 6 is entirely backward looking, meaning that current consumption depends only on current price, lagged consumption, the marginal utility of wealth (which is one of the determinants of the current price coefficient), and current events. Current consumption is independent of both future consumption, C_{t+1}, and future events, e_{t+1}. Because of these distinctions, myopic models and rational models have different implications regarding responses to future changes. In particular, rational addicts increase their current consumption when future prices are expected to fall, but myopic addicts do not.[7]

The parameters of equation 6 also allow one to compute the long- and short-run responses of consumption to a permanent decline in price. The long-run response is the effect on consumption of a change in price in all periods. The short-run response is the effect on consumption of an unanticipated change in price in the current period and all future periods. (That is, past consumption is held constant). These effects are obtained by solving the second-order difference equation in equation 6. This results in a reduced-form

[7] *The above proposition is in sharp contrast to the claim by Spinnewyn (1981) that myopic behavior is empirically indistinguishable from rational behavior in the context of models of habit formation. Pashardes (1986) takes the same position as Becker, Murphy, and I take. He provides an empirical test of myopic versus forward-looking behavior in the context of the Almost Ideal Demand System. His procedures, however, are extremely difficult for me to follow. I urge the interested reader to consult Parshades' paper.*

demand function in which current consumption depends on current price and on *all* past and future prices. The long-run price response also can be obtained by equating C_t, C_{t+1}, and C_{t-1} to obtain the steady-state value of consumption:

$$(\partial C/\partial P) = [\alpha_1/(1 - \alpha - \beta\alpha)] \qquad (7)$$

where the denominator of equation 7 must be positive for a stable steady state. Clearly, the long-run price effect rises in absolute value as the degree of addiction (α) rises.[8] Since changes in α need not change C, the long-run price elasticity (equation 7 multiplied by the ratio of P to C) also tends to rise in absolute value as the degree of addiction rises.

In addition, the solution to the second-order difference equation in equation 6 yields temporary past and future price effects. Both must be negative according to the theory of rational addiction. The temporary past price effect is defined as the impact of a change in price in period t-1, with all other prices held constant. Similarly, the temporary future price effect is defined as the impact of a change in price in period t+1, with all other prices held constant.[9]

As previously indicated, a rational addiction model predicts that the long-run price effect or elasticity should exceed the short-run price effect or elasticity. This same prediction is contained in a myopic model of addiction. But the latter model predicts that the future price effect should be zero, while the former model predicts that it should be negative. Moreover, the short- and long-run price effects that emerge from a myopic specification of equation 6 are biased if the coefficient of future consumption is significant.

So far I have dealt with a simple version of the rational addiction model in which the rate of depreciation on the addictive stock equals one. In a more general version of the model, this rate is smaller than one. Becker et al. (1991*a*) show that the following demand function emerges in this model:

$$\begin{aligned} C_t &= \theta C_{t-1} + \beta\theta C_{t+1} \\ &\quad + \theta_1[1 + (1-\delta)^2\beta]P_t \\ &\quad - \theta_1(1-\delta)P_{t-1} - \beta\theta_1(1-\delta)P_{t+1} \end{aligned} \qquad (8)$$

In this equation the unmeasured life-cycle variables that affect utility are suppressed, θ is positive, and θ_1 is negative.

This equation is a generalization of equation 6. The major difference is that past and future prices enter in addition to current price and past and future consumption. Nevertheless, the basic estimation strategy remains the same. Since the relevant unobservable variables are likely to be correlated over time, instrumental

[8] *For a more detailed proof, see Becker et al. (1991b). The proposition in the text does not necessarily imply that addictive goods are more responsive to price than are nonaddictive goods. Let* α_{cc} *and* α_{yy} *be the coefficients of* C^2 *and* Y^2*, respectively, in the quadratic utility function. The long-run price responsiveness of C exceeds the short- or long-run price responsiveness of Y (both are the same since Y is nonaddictive) if* α_{cc} *equals* α_{yy}*. But if* α_{cc} *exceeds* α_{yy} *in absolute value, the result is ambiguous.*

[9] *See Becker and Murphy (1988) and Becker et al. (1991a) for a discussion of the difference between temporary anticipated cross-price effects and temporary unanticipated cross-price effects.*

variables for C_{t-1} and C_{t+1} are required to obtain consistent estimates of the parameters. Longer lags of the price variable are logical candidates for instruments. Additional empirical leverage is available from the condition that the discount factor, β, equals both the ratio of the coefficient of C_{t+1} to that of C_{t-1} and the ratio of the coefficient of P_{t+1} to that of P_{t-1}. The general model reduces to the simple model when the rate of depreciation on the addictive stock equals one. Clearly, the coefficients of past and future price in equation 8 are zero if δ equals one. In that case equations 6 and 8 are identical.

Becker et al. also derive an equation for myopic behavior that is similar to the myopic demand function described above except that past price also enters on the right-hand side. Once again, myopic behavior is nested within the rational behavior: myopic behavior implies zero coefficients in equation 8 for future price and future consumption.

For the estimated version of equation 8, rational addiction implies positive coefficients on past and future consumption, positive coefficients on past and future prices, and a negative coefficient on current price. Positive coefficients on past and future prices may seem odd, given that past and future consumption are complementary with current consumption when behavior is addictive. However, controlling for past consumption eliminates the channel through which past prices affect current consumption. But the only way that past consumption stays fixed when past prices are higher would be for another force to offset higher past prices by raising the past stock of consumption capital. Since this higher value of the stock continues into the present period (reduced only by depreciation), current consumption must be higher when past prices are higher. This also explains why past and future prices were not in the estimating equation of the simple model. That model implies a depreciation rate of one, so that any larger past stock is eliminated entirely by depreciation.

To obtain the short-run price effect, the long-run price effect, and the temporary past and future price effects defined above, one solves the second-order difference equation in equation 8. Note that the past and future price effects are negative given "adjacent complementarity." Becker and Murphy (1988) define this concept as the necessary and sufficient condition for an increase in the stock of an addictive good to increase current consumption, or for reinforcement to be present. If the depreciation rate on the addictive stock equals one or if consumers are myopic, this condition is satisfied if an increase in the stock raises the marginal utility of current consumption ($\partial^2 U/\partial C \partial S = U_{CS} > 0$). But if consumers are rational and the depreciation rate is smaller than one, Becker and Murphy (1988) show that a positive value of U_{CS} is necessary but not sufficient for reinforcement. The necessary and sufficient condition is

$$(r + 2\delta)U_{CS} > -U_{SS} \qquad (9)$$

where $r = (1-\beta)/\beta$ is the rate of time preference for the present and U_{SS} is negative by concavity. This condition underscores

that rational utility maximizers consider the future harmful consequences of their current behavior. Reinforcement for them requires that the positive effect of an increase in S on the marginal utility of current consumption exceeds the negative effect of higher S on the future harm from higher current consumption.

Becker et al. (1991*b*) develop some theoretical extensions of the rational addiction model. We note that the total cost of addictive goods to consumers equals the sum of the good's price and the money value of the future adverse effects, such as the negative effects on earnings and health of smoking, heavy drinking, or heavy dependence on crack. We also note that future costs tend to be less important to poorer, less educated, and younger consumers because they generally place a smaller monetary value on health and other harmful future effects than richer, more educated, and older consumers who have higher wage rates. We note finally that the poor, youths, and the less educated are likely to have lower time-discount factors (higher rates of time preference for the present) than the rich, adults, and the more educated.[10] We then show that the poor, youths, and the less educated are more sensitive to changes in money prices of addictive goods, whereas the middle or upper income class, adults, and the more educated respond more to changes in the perceived or actual harmful consequences that take place in the future.

The final notable component of Becker and Murphy's (1988) treatment of rational addiction pertains to their analysis of unstable steady states. A steady state is one in which annual consumption of an addictive good is exactly equal to depreciation of the addictive stock, so that the stock remains constant over time. An unstable steady state is one in which a price change causes persons to start or stop consuming the addictive good or to change their consumption by a very large amount. Becker and Murphy show that unstable steady states are possible in the context of their rational addiction model. They then point out, "With two steady states, relatively few persons consistently consume small quantities of addictive goods. Consumption diverges from the unstable steady state toward zero or toward the sizable steady-state level. Therefore, goods that are highly addictive to most people tend to have a bimodal distribution of consumption, with one mode located near abstention" (Becker and Murphy 1988, p. 683).

The preceding ideas provide a theoretical framework to study the decisions to start, stop, or restart consuming an addictive good. Prices are important in these decisions, but they also depend on the occurrence of stressful life events such as unemployment or divorce. A novel implication of the Becker-Murphy model is that the probability of consuming an addictive good in the current period and the quantity of current consumption, given participation, will rise if a change in status from married or divorced or employed to unemployed is anticipated in the not-too-distant future.

[10] *This proposition is suggested by the theoretical analysis in Becker (1990).*

Becker and Murphy (1988) and Becker et al. (1991*b*) are not the only examples of recent contributions to the theory of rational addiction. Barthold and Hochman (1988) and Michaels (1988) also have contributed to this literature. But from the point of view of empirical research, the model by Becker and Murphy appears to be the most relevant.[11]

EMPIRICAL APPLICATIONS OF RATIONAL ADDICTION: CIGARETTES, GAMBLING, AND LEISURE

Becker, Murphy, and I (Becker et al. 1991*a*) fit models of rational addiction to cigarettes to a U.S. time series of State cross-sections for the period 1955–85. We focus on the simple model in which the rate of depreciation on the addictive stock equals one (see equation 6), although we also obtain estimates of the general model in which the rate of depreciation on the addictive stock is less than one (see equation 8). Structural demand functions are fitted by two-stage least squares with past and future consumption treated as endogenous right-hand side variables. Past and future cigarette prices are used as instruments in all specifications of equation 6.[12] Current, past, and future State cigarette excise tax rates are added to the set of instruments in some versions. This is because consumers may have more knowledge about taxes, especially future taxes, than about future prices.

The estimated coefficient of current price in equation 6 is negative and statistically significant, while the estimated coefficients of past and future consumption are positive and significant. These results imply both that cigarette smoking is addictive and that consumers are rational rather than myopic. We find a sizable long-run price elasticity of demand of approximately –0.77, which is almost twice as large as the short-run price elasticity of –0.40. Smoking in different years appear to be complements: Cigarette consumption in any year is lower when both future prices and past prices are higher.

Clearly, the estimates indicate that cigarettes are addictive, and that past and future changes significantly impact current consumption. This evidence is inconsistent with the hypothesis that cigarette consumers are myopic. Still, the estimates may not be fully consistent with rational addic-

[11] *Given my close association with Becker and Murphy, the above evaluation may be biased. I urge other researchers to evaluate the potential utility of alternative economic models of addictive behavior. Michaels (1988) makes the important point that an antidrinking campaign that promotes the productivity of activities other than alcohol abuse will be more successful than one that focuses solely on the harmful consequences of abuse. This is a testable proposition. Barthold and Hochman (1988) predict that free outpatient treatment programs would be less successful than inpatient programs, which is testable in the context of treatment research.*

[12] *We treat price as exogenous. This implies that the cigarette market is competitive and that the average or marginal cost of producing cigarettes is constant. Using the same basic data that we employ, Sumner (1981, p. 1017) concludes, "the effect of monopoly power as a source of price variation in the [cigarette] industry is small compared with tax rate differences and other cost differences over time and space." In the context of a nonaddictive model, Porter (1986) reports little difference between cigarette market demand functions that treat price as exogenous and demand functions that treat price as endogenous.*

tion. The point estimates of the discount factor (β) are implausibly low—the ratio of the coefficient of future consumption to the coefficient of past consumption ranges from 0.31 to 0.64. These discount factors correspond to interest rates ranging from 56.3 to 222.6 percent. However, since future prices may not be fully anticipated by consumers, estimates based on the assumption of perfect foresight may overstate the variability of expected future prices. Due to an errors-in-variables bias, this may lead to an understatement of the effect of future prices on current consumption. Therefore, uncertainty about future prices could explain the implausibly high interest rates implied by our estimates.

To explore the preceding issue in more detail,[13] we excluded the one-period lead value of the price and the one-period lead value of the excise tax from the set of instruments used to obtain the demand functions. The resulting estimates are very different from the ones that include these variables as instruments. The new estimates still support the rational addiction model because the coefficient of future consumption is positive and significant. But the results provide little support for the hypothesis that the implied discount factors are low because future prices and taxes are inappropriate instruments. Indeed, we tested this hypothesis directly by adding the future price and the future tax to the set of second-stage regressors while continuing to exclude these variables as instruments in the first stage. If the correct set of instruments excludes future variables, then these variables should be insignificant as a set in the second stage. This hypothesis is rejected.

As a further refinement, we imposed six alternative discount factors ranging from 0.70 to 0.95 (interest rates ranging from 5.3 to 42.9 percent). That is, we constrained the coefficient of future consumption to equal the discount factor multiplied by the coefficient of past consumption. We imposed this constraint both in the specifications that include the future price and the future tax as instruments and in the specifications that exclude these variables as instruments. Regardless of the discount factor imposed, the long-run price elasticities are very similar to each other and to those in the unconstrained models that use the future price and the future tax as instruments. The same comment applies to the short-run price elasticities. Moreover, the constrained specifications that employ the future price and the future tax as instruments yield elasticities that are almost identical to the constrained specifications that exclude these two instruments. We do, however, obtain one piece of evidence in favor of the use of future prices and taxes as instruments. When these variables are used as instruments in obtaining the constrained model, the restriction is not significant at the 1-percent level in 14 out of 18 cases, and it is not significant at the 5-percent level in 11 out of 18 cases based on a Lagrange Multiplier (LM) test. On the other hand, when future variables are not used as instruments, the restriction is

[13] *I discuss the issue of the low discount factor and measurement error in future prices in detail because this has been the main criticism of the Becker et al. (1991a) study.*

significant in every case at any conventional level of confidence. Stated differently, the imposed discount factors are valid in most specifications that include future prices and taxes as instruments, but they are not valid in the specifications that omit these instruments.

We view these results as rather strong evidence in support of the rational addiction model and the use of future prices and taxes as instruments to estimate the model. In reaching this conclusion, we note that consumers are likely to have a good deal of information concerning the State-specific future price of cigarettes because this price depends to a large extent on the future State excise tax rate on cigarettes. Excise tax hikes are announced in advance and receive a good deal of publicity as a result of the legislative process. Moreover, most States raise their excise tax rates in response to revenue shortfalls. Hence, it is plausible that tax hikes are anticipated even before bills to do so are introduced in State legislatures.

We also note that there are two main predictions of the rational addiction model: (1) a positive and significant coefficient of future consumption and (2) a future consumption coefficient that is smaller than the coefficient of past consumption. Both predictions are confirmed by the unrestricted estimates. As long as the implied discount factor is positive and smaller than one, the model does not stand or fall on the specific discount factor that is obtained. The results suggest that the data are not rich enough to pin down the discount factor with precision. This is not surprising. Estimates of consumer discount factors from studies of aggregate consumption, the consumption of specific goods, or the consumption of leisure over time vary considerably. Some of these estimates imply extremely high interest rates, while others imply very low and even negative interest rates (see Becker et al. 1991*a* for references to these studies). Nevertheless, it is reassuring that our estimates of the basic parameters of the model are not sensitive to the choice of alternative discount factors. Moreover, in the specifications with future price and tax as instruments, we cannot reject the hypothesis (at the 1 percent level) that the discount factor is as high as 0.90 or 0.95 in four of six cases. Finally, when we compensate for the narrow set of instruments that results from the deletion of future variables by imposing a discount factor, the estimates of short-run and long-run price elasticities are not sensitive to the instruments used to obtain them.

Even though our estimates indicate that the equilibrium aggregate amount of smoking is stable, unstable steady states may greatly affect the overall response of cigarette smoking to changes in cigarette prices. This conclusion is supported by the evidence that almost all the effect of higher prices on teenage smoking and more than half the effect on adult smokers are due to a decline in the number of smokers (Lewit et al. 1981; Lewit and Coate 1982). The possible role of unstable steady states is inferred from the fact that the large decline in the number of smokers is not due entirely to heterogeneity in the amount smoked, whereby people who smoke only a few cigarettes per day stop

smoking when cigarette prices increase. Data from the second National Health and Nutrition Examination Survey reveal that 60 percent of persons who had stopped smoking for less than 1 year smoked at least one pack of cigarettes per day during the last year in which they smoked. The same survey indicates that 68 percent of persons who had stopped smoking for 1 or more years smoked at least one pack a day during their period of maximum consumption. Even a modest increase in cigarette prices could induce smokers who happen to be near unstable steady states to cease (see the analysis in Becker and Murphy 1988). The existence of unstable steady states means that some people who smoke a lot will be the "marginal" smokers with respect to changes in cigarette prices and other variables.

Chaloupka (1991) provides further evidence in support of the rational model of cigarette addiction in a micro data set using data from the second National Health and Nutrition Examination Survey. Using measures of cigarette consumption in three adjacent periods, he fits demand functions similar to those in the Becker et al. 1991*a* study. He reports positive and significant future and past consumption coefficients and a short-run price elasticity (–0.20) that is less than one-half of the long-run price elasticity of –0.45. He also finds that smoking by the less educated is considerably more responsive to changes in cigarette prices than is smoking by the more educated; a similar result has been obtained by Townsend (1987) with British data. Lewit et al. (1981) and Lewit and Coate (1982) report that youths respond more to cigarette prices than do adults. By contrast, the information that began to emerge in the early 1960's about the harmful long-run effects of smoking has had a much greater effect on smoking by the rich and more educated than by the poor and less educated (Farrell and Fuchs 1982 for the United States; Townsend 1987 for Britain).[14]

Keeler et al. (1991) fail to confirm the rational addiction model in a monthly time-series study of the demand for cigarettes in California during the period 1980–90. They do not, however, estimate the structural version of the Becker et al. 1991*a* model. Instead, they regress cigarette consumption on the current price, three lags of the price, and three leads of the price. As Becker et al. (1991*a*) point out, since future and past prices have direct effects on future and past consumption, future and past prices indirectly impact current consumption when the current price is held fixed. Both future and past prices negatively affect the consumption of rational addicts. Keeler et al. report positive and insignificant past and future price coefficients, however.

[14] *These results are consistent with the proposition in the third section of this paper that the poor, youths, and the less educated are more sensitive to changes in monetary prices of addictive goods, whereas the middle or upper income class, adults, and the more educated respond more to changes in the perceived or actual harmful consequences that take place in the future. The interaction between age and price also may be traced to synergies between peer pressure and addiction discussed by Becker (1991). Note that Wasserman et al. (1991) fail to replicate the interaction between age and price in the demand for cigarettes using various National Health Interview Surveys in the 1970's and 1980's.*

Becker et al. (1991*a*) argue that the coefficients that emerge from an equation like the one that Keeler et al. fit are not consistent estimates of the true cross-price effects due to the omission of other prices. That is, the equation they fit is a truncated version of the reduced-form demand function for cigarettes that results from solving the second-order difference equation 6. All past and future prices enter this function. Becker, Murphy, and I fit a truncated version of it with one lead and one lag of price as a preliminary check of our model. We obtain negative and significant cross-price coefficients. The findings of Keeler et al. may be due to the inclusion of too many leads and lags and to multicollinearity among monthly current, past, and future prices for a single State.

Mobilia (1990) applies the rational addiction framework to the demand for gambling at horse-racing tracks. Her data consist of a U.S. time series of horse track cross-sections for the period 1950–86 (tracks over time are the units of observation). She measures consumption by the real amount bet per attendant (handle per attendant) and price by the takeout rate (the fraction of the total amount bet that is retained by the track). Her findings are similar to those in the rational addictive studies of cigarettes by Becker et al. (1991*a*) and Chaloupka (1991). The long-run price elasticity of demand for gambling at horse tracks equals −0.7, which is more than twice as large as the short-run elasticity of −0.3. Moreover, an increase in the current takeout rate lowers handle per attendant in past and future years.

Mobilia also considers attendance per capita as a dependent variable. She finds no support in favor of the rational addiction model with this outcome. Her results are plausible because handle per attendant is a much better measure of addiction to gambling than attendance per capita. That is, a considerable number of people attend horse-racing tracks for entertainment and place very modest bets. These persons are not addicted to gambling. Moreover, Mobilia cites data that indicate that 20 percent of bettors account for 80 percent of all bets.

Hotz et al. (1988) and Bover (1991) apply variants of rational addiction models to the demand for leisure time or the supply of hours of work over the lifecycle. Both studies use panel data on males in the University of Michigan's Panel Study of Income Dynamics. Both report evidence of rational addiction in the sense that current hours of work are positively related to past and future hours of work.

In this section and in the third section of this paper, I have reviewed empirical studies on the demand for addictive goods both in the context of myopic models of addiction and in the context of rational models of addiction. Research on the supply side of the market for these goods is much more limited. Becker et al. (1991*a*) briefly consider the optimal pricing policy of a rational monopolistic producer of an addictive good. We show that the monopolist will set current price such that marginal revenue is less than marginal cost since future profits are higher when current consumption is larger (current price is smaller) because greater current

consumption raises future consumption. We also show that current price and current profits can rise in the face of a drop in demand due, for example, to an increase in information about the harmful effects of cigarette smoking. Current price rises because a decline in future demand reduces the gains from lowering price now. In turn, current profits rise because the monopolist previously was pricing below the current period profit-maximizing point. These trends in demand, price, and profits have characterized the cigarette industry since 1981.

Finally, we (Becker et al. 1991*a*) show that a monopolistic producer may raise current price in response to an increase in future costs. A higher Federal excise tax on cigarettes was widely expected to go into effect at the beginning of 1983. Cigarette prices increased sharply not only in 1983 but also prior to the tax increase during 1982. In the context of rational addiction, this occurred because the higher future cigarette tax reduced future demand and hence the gain from lowering current price.

Showalter (1991) develops a more elaborate model of the optimal pricing policy of a monopolistic producer of an addictive good. He also considers the behavior of consumers of the good. He treats four cases: myopic monopolist–myopic consumer, rational monopolist–myopic consumer, myopic monopolist–rational consumer, and rational monopolist–rational consumer. These cases are subjected to empirical testing using the same data set employed by Becker et al. (1991*a*). Showalter's results indicate that the rational monopolist–rational consumer model provides the best fit, although he treats past and future consumption as exogenous in estimating the demand function.

RATIONAL ADDICTION AND THE DEMAND FOR EXCESSIVE ALCOHOL CONSUMPTION

In the previous section I emphasized empirical applications of the rational addiction model to the demand for cigarettes. Attempts to apply similar models to the demand for alcohol consumption are inherently more difficult for two reasons. First, alcohol consumption is not nearly as addictive as cigarette smoking since many persons consume relatively small quantities of alcohol but not cigarettes. Stated differently, the distribution of alcohol consumption is more continuous than the bimodal distribution that Becker and Murphy (1988) show is likely to characterize consumption of an addictive good.

Second and in light of the factor just mentioned, applications of rational addiction to alcohol consumption must pay attention to the behavior of heavy drinkers. But then one confronts the stereotype of the alcoholic who is completely unresponsive to price. One also confronts a second stereotype of the heavy drinking nonalcoholic—a person who drinks heavily from time to time but whose annual consumption is smaller than that of an alcoholic. As in the case of alcoholics, the conventional wisdom is that heavy drinking occasions and consumption by heavy drinkers are not sensitive to price.

If the conventional wisdom with respect to the behavior of alcoholics and heavy drinkers were correct, it would be difficult to justify the application of rational addiction models to alcohol consumption. Therefore, I begin this section by reviewing evidence that suggests that this is not the case. In a clinical experiment, Bigelow and Liebson (1972) used two male, skid row, chronic alcoholic volunteers residing in Baltimore to show that alcohol-dependent persons reduce their alcohol consumption as a function of beverage costs. Babor et al. (1978) recruited 20 adult male volunteers with a prior history of casual drinking and 14 adult male volunteers with a prior history of heavy drinking to conduct an experimental study of price reductions during afternoon "happy hours" in Boston. Approximately half the subjects in each category could purchase alcohol under a single-price condition (50 cents per drink), while a matched group was given a price reduction daily (25 cents per drink) during a 3-hour period in the afternoon. The afternoon price reductions significantly increased alcohol consumption by both casual and heavy drinkers. Reinstatement of the standard purchase price reduced consumption by both groups.

Kendell et al. (1983) examined alcohol consumption by heavy and moderate drinkers before and after Scotland increased its excise tax on alcoholic beverages in March 1981. They used a sample of 463 residents of the Lothian region (Edinburgh and its hinterland). Alcohol consumption and its associated adverse effects (for example, getting into a fight due to drinking, being in a road accident after drinking) fell proportionately more among heavy drinkers than among other drinkers.

Although the above evidence is impressive, the studies cited should be interpreted with some caution because of the limited and nonrepresentative nature of the samples employed. Clearly, it would be misleading to generalize their results to the population at large. But these studies highlight the potential payoffs of research that is directed toward obtaining refined estimates of demand functions for alcohol use and abuse by persons of different ages.

More definitive evidence with regard to the behavior of heavy drinkers is contained in a well-known study by Cook and Tauchen (1982). They examine variations in death rates from cirrhosis of the liver (a standard measure of excessive alcohol use) as well as variations in per capita consumption of distilled spirits in a time series of license (open) State cross sections for 1962–77. They find that the State excise tax rate on distilled spirits has a negative and statistically significant effect on the cirrhosis death rate. Moreover, a $1 increase in the State excise tax rate lowers the death rate by approximately the same percentage (between 5.4 and 10.8 percent) as it lowers per capita consumption (7.2 percent). Cook and Tauchen conclude that "liquor consumption, including consumption of heavy drinkers, is quite responsive to price" (1982, p. 387).

My research on youth alcohol use in micro data with Coate and Arluck

(Grossman et al. 1987; Coate and Grossman 1988), and research by Kenkel (in press) and by Manning et al. (1991) on alcohol use by all segments of the population in similar data, also highlight the importance of price. My research deals with youths ages 16 through 21 in the first National Health and Nutrition Examination Survey, conducted between 1971 and 1975 (Grossman et al. 1987), and in the second National Health and Nutrition Examination Survey, conducted between 1976 and 1980 (Coate and Grossman 1988). We find that the use of alcohol by youths is inversely related to the prices of alcoholic beverages in both data sets. The beverage-specific price is a particularly important determinant of beer consumption. This is a key result because beer is the drink of choice among youths who consume alcoholic beverages. The negative and statistically significant price effect is by no means limited to reductions in the fraction of youths who consume beer infrequently (less than once a week). Instead, the fractions of youths who consume beer fairly frequently (one to three times a week) and frequently (four to seven times a week) fall more in absolute or percentage terms than does the fraction of infrequent drinkers when price rises. Along the same lines, the fractions of fairly heavy (three to five cans on a typical drinking day) and heavy (six or more cans on a typical drinking day) youthful beer drinkers decline more in absolute or percentage terms than the fraction of light (one to two cans on a typical drinking day) drinkers in response to price increases. Kenkel's (in press) outcome measure is the number of days in the past year on which a person consumed five or more drinks in the 1985 National Health Interview Survey. He reports that this measure is inversely related to price among adults and youths. Manning et al. (1991) report a similar finding for this outcome and for average daily consumption of ethanol in the 1983 National Health Interview Survey.[15]

The studies by Cook and Tauchen (1982), Grossman et al. (1987), Coate and Grossman (1988), Kenkel (in press), and Manning et al. (1991) are conducted in the context of a standard model of consumer behavior rather than in the context of a model of rational addiction. That is, no attempts are made to estimate a structural demand function in which current consumption depends on past and future consumption, both of which are endogenous, as well as on current price. Stated somewhat differently, the addiction model generates a reduced-form demand function in which current consumption depends on current price and on past and future prices. The studies at issue fit reduced form demand functions with past and future prices omitted. In the context of the addictive model, the price coefficients that they obtain may be viewed as biased estimates of long-run price effects. To those who do not accept the assumptions of the rational addictive model or the empirical evidence in support of it (perhaps because they think that the

[15] *Manning et al. (1991) do not present separate estimates for youths and adults.*

econometric tests of it are too taxing on the data), the price coefficients in the research at issue are unbiased. Regardless of one's views on this issue, the main message of these studies is that it is a mistake to assume that incidence of heavy drinking and consumption by heavy drinkers are not sensitive to price.

Currently, Becker, Chaloupka, Murphy, and I (Grossman et al. in progress) are applying the rational addiction model to alcohol outcome measures in a time series of State cross sections for 1962–84. These measures are per capita consumption of distilled spirits, total per capita ethanol consumption (an aggregation of the ethanol in beer, wine, and spirits), and the age-adjusted cirrhosis mortality rate of persons age 30 and over. The price of alcohol is given by a Divisia index of the prices of the three leading brands of spirits in the period 1958–84. These are Seagram's 7-Crown (a blended whiskey), Smirnoff (vodka), and Bacardi (rum, for 1958–64 the brand of rum is Carioca). We begin the price series in 1958 to experiment with alternative lags of price as instrumental variables. Prices for 1973–84 were taken from the Distilled Spirits Council of the United States (DISCUS), and prices prior to 1973 were taken from the Jobson Publishing Company (formerly Gavin-Jobson). These sources stopped reporting prices after 1984.

Unlike Cook and Tauchen (1982), we include both license and monopoly (control) States in our database. We use the real State excise tax on beer as an additional instrument in the same manner as Becker et al. (1991*a*) use the real State excise tax on cigarettes as an additional instrument. This is possible because beer can be sold privately in monopoly States. In the period at issue, the simple correlation coefficient between the real price of alcohol [the Divisia index divided by the (Consumer Price Index, CPI)] and the real beer tax (the sum of the State and Federal tax divided by the CPI) equals 0.62.

When per capita consumption of distilled spirits or total per capita consumption of alcohol is used as the dependent variable, we do not find much evidence in support of the rational addiction model. This is not surprising since alcohol consumption is not nearly as addictive as cigarette smoking, as pointed out above. That is, the results indicate that the aggregate data are dominated by the behavior of light and moderate drinkers.

More promising results are obtained when the age-adjusted cirrhosis mortality rate at time $t+1$ (M_{t+1}) is used to develop a measure of excessive alcohol consumption at that time. We begin with the assumption that this rate is a function of excessive alcohol consumption in each previous period (C_i, $i = 1,2,\ldots,t$) with exponentially declining weights:

$$M_{t+1} = \sum_{i=1}^{t} \gamma(1-\delta)^{t-i} C_i \qquad (10)$$

Here, γ is the positive effect of an increase in C_t on M_{t+1}, and δ is the constant rate at which the adverse health effects of heavy drinking at time t depreciate. Since equation 10 holds for M_t as well as for M_{t+1},

$$M_{t+1} = \gamma C_t + (1-\delta)M_t \qquad (11)$$

Equation 8 specifies a demand function for C_t in the context of the general model of rational addiction (the model in which the rate of depreciation on the addictive stock is less than one). Substitute this demand function into equation 11 and make use of the fact that $\gamma C_{t-1} = M_t - (1-\delta)M_{t-1}$ and $\gamma C_{t+1} = M_{t+2} - (1-\delta)M_{t+1}$ to obtain

$$M_{t+1} - (1-\delta)M_t = \theta[M_t - (1-\delta)M_{t-1}] + \beta\theta[M_{t+2} - (1-\delta)M_{t+1}] + \gamma\theta_1[1 + (1-\delta)^2\beta]P_t - \gamma\theta_1(1-\delta)P_{t-1} - \beta\gamma\theta_1(1-\delta)P_{t+1} \quad (12)$$

Several comments are in order concerning equation 12. First, if δ were known, consistent estimates of its parameters could be obtained by two-stage least squares with $M_t - (1-\delta)M_{t-1}$ and $M_{t+2} - (1-\delta)M_{t+1}$ treated as endogenous. Since this rate is not known, we fit equation 12 for assumed values of δ ranging between 0.1 and 0.9.

Second, unlike in the simple model of rational addiction with a rate of depreciation on the addictive stock equal to one, the one-period lead and lag of price cannot be used to identify the model. That is, the members of the set of identifying instrumental variables become smaller, and the general model is more "taxing" on the data than the simple model. We address this issue by imposing the three constraints implied by the model. These are (1) the coefficient of $M_{t+2} - (1-\delta)M_{t+1}$ equals the coefficient of $M_t - (1-\delta)M_{t-1}$ multiplied by the discount factor (β); (2) the coefficient of future price (P_{t+1}) equals the coefficient of past price (P_{t-1}) multiplied by β; and (3) the ratio of the coefficient of current price (P_t) to the coefficient of past price equals $-[1 + (1-\delta)^2\beta]/(1-\delta)$. For a given value of δ, the constraints are imposed for alternative discount factors ranging from 0.8 to 0.95. We test the significance of each restriction using a Lagrange Multiplier test.

Very preliminary estimates of the model are presented in table 1. Each of the two specifications in the table sets δ equal to 0.5 and β equal to 0.95. Both models include dichotomous variables as regressors for each year in the sample period except one. The first model in the table also includes real per capita income; the percentage of the population age 25 and older with at least a high school education; the percentage of the female population ages 25–34 that are divorced; the unemployment rate; and the percentages of the population that are Mormon, Southern Baptist, Catholic, and Protestant (other than Mormon and Southern Baptist), respectively.[16] The second replaces all the demographic variables except income with dichotomous variables for each State except one. In both models the two-period lagged value of the price, the current beer tax, the first two lags of the beer tax, the first lead of the beer tax, and the first three lags of income and the unemployment rate are used as first-stage but not second-stage regressors.[17] The price and consumption coefficients in table 1 have the correct

[16] *The definitions, sources, and roles of the demographic variables are discussed in detail by Becker et al. (1991a).*

[17] *Of course, all the exogenous variables in the second stage are employed as regressors in the first stage.*

TABLE 1

Restricted two-stage least-squares demand functions for excessive alcohol consumption

	(1)	(2)
$M_t - 0.5M_{t-1}$	0.432	0.427
	(8.20)	(4.58)
$M_{t+2} - 0.5M_{t+1}$	0.411	0.405
	(8.20)	(4.58)
P_{t-1}	0.070	0.048
	(2.04)	(1.37)
P_t	–0.172	–0.118
	(2.04)	(–1.37)
P_{t+1}	0.066	0.045
	(2.04)	(1.37)
Marginal significance level *		
Restriction 1	0.957	0.868
Restriction 2	0.416	0.142
Restriction 3	0.319	0.100
R-squared	0.566	0.674
N	1,000	1,000

Note: Dependent variable = $M_{t+1} - 0.5M_t$. Asymptotic t-statistics in absolute value in parentheses. Intercepts not shown. $M_t - 0.5M_{t-1}$ and $M_{t+2} - 0.5M_{t+1}$ are endogenous. $\beta = 0.95$ and $\delta = 0.5$. See text for other exogenous variables in each model.
** Based on a Lagrange Multiplier (LM) test.*

signs and are statistically significant, with the exception that the price coefficients are significant at the 10-percent level on a one-tailed test but not at the 5-percent level in model 2. None of the three restrictions is significant at the 1-percent or the 5-percent levels of significance. Stated differently, the three restrictions, which are based on a discount factor of 0.95 and a depreciation rate of 0.5, are valid.

The first model in the table implies a substantial long-run price elasticity of demand of –1.28 for excessive alcohol consumption. The corresponding short-run elasticity of –1.07 is approximately 20 percent smaller than the long-run elasticity. An increase in current price leads to reductions in consumption next period and last period, although the elasticities at issue (–0.04 in either case) are relatively small. In model 2 the long-run elasticity is –0.83, the short-run elasticity is –0.72, and the temporary past and future price elasticity both equal –0.02. Taken together, these results imply that excessive alcohol consumption is addictive and that consumers are farsighted rather than myopic.

Currently, we are evaluating estimates such as those in table 1 in more detail. Issues that still must be resolved include the selection of an appropriate depreciation rate and a test of the goodness of fit of the constrained rational addiction model versus the constrained myopic addiction model. These are nonnested specifications, so that a variant of the Davidson-McKinnon J Test (Davidson and McKinnon 1981) may be in order.[18] Given the very preliminary nature of our results, we view them as suggestive rather than definitive evidence in favor of the rational addiction model. We note, however, that they do indicate that the demand for excessive alcohol consump-

[18] *A variant may be required since our models are estimated by two-stage least-squares.*

tion may be quite responsive to price. They are also similar to results obtained by Mobilia (1990) in her study of gambling at horse-racing tracks that I discussed earlier in the fourth section. She finds that the amount bet per attendant is characterized by rational addiction, but attendance per capita is not characterized by this phenomenon. We find that per capita alcohol consumption is not characterized by rational addiction in data aggregated to the State level, but rational addiction may well characterize excessive alcohol consumption.

Our results can be compared to those in a study by Heien and Pompelli (1987). They find no impact of distilled spirits prices on cirrhosis mortality rates in a time series of State cross sections for the period 1968–77. They assume, however, that mortality is a perfect proxy for heavy consumption instead of adopting the more flexible and general distributed lag specification given by our equation 10. Moreover, they ignore the roles of past and future consumption and price variables in the demand function. Their findings suggest that the exclusion of these variables may give a misleading picture of the magnitude of the long-run price effect.[19]

TOPICS FOR CONTINUING RESEARCH ATTENTION

Micro panel data are the ideal source to estimate myopic and rational addiction models of alcohol consumption and excessive consumption. In addition, such data are required to explore the decisions to start, stop, and restart heavy drinking using insights provided by Becker and Murphy's (1988) theory of rational addiction. I place this topic at the top of an agenda for continuing and future research attention.

The research just mentioned requires measures of an individual's alcohol consumption in at least three periods (years) of his/her lifecycle. It also requires consistent measures of alcoholic beverage prices both over time and in the cross-section. In 1988 Jon Nelson uncovered the beer, wine, and distilled spirits prices in the *Inter-City Cost of Living Index*, published quarterly by the American Chamber of Commerce Research Association (ACCRA) for approximately 250 cities since 1968 (see Nelson 1990 for the first paper that resulted from the discovery of the ACCRA alcoholic beverage prices). Until then I was pessimistic about future research with micro data in light of DISCUS's decision to cease publishing spirits prices and in the absence of regular sources of beer and wine prices.[20] To be sure, the State beer tax is a good proxy for the price of beer in all States since beer can be sold privately in monopoly States. But State excise tax rates on wine and liquor are poor proxies for the prices of wine and beer in monopoly States because such States derive most of their revenue from the sale of wine and liquor from

[19] *It appears as if the price variable used by Heien and Pompelli (1987) varies by State but not by year, although this is ambiguous in their paper.*

[20] *My research on the demand for beer by youths (Grossman et al. 1987; Coate and Grossman 1988) relied on special one-time surveys of beer prices.*

price markups rather than from excise taxes. Moreover, attempts to fit rational addiction models by Becker et al. (1991*a*) and by Grossman et al. (in progress) have used both prices and taxes as instruments.

The ACCRA has collected spirits prices since 1968 and beer and wine prices since 1982. Nelson (1990) employs these prices to fit demand functions for beer, wine, and distilled spirits in a cross-section of States for 1982. His estimates of price elasticities are similar to those in previous research which relied on other sources for prices. Kenkel (in press) and Manning et al. (1991) use ACCRA prices in their micro demand studies of the demand for alcohol in the 1985 and the 1983 National Health Interview Surveys, whose results I described in the fifth section of this paper. Phelps and Leung (in progress) are conducting research on a similar topic with these prices in the 1983 National Health Interview Survey. Becker, Chaloupka, Murphy, and I have been able to replicate my published results on teenage alcohol use in the 1982 and 1989 cross-sectional surveys of high school seniors conducted by the Institute for Social Research of the University of Michigan as part of an ongoing national research program entitled "Monitoring the Future." This research also relies on the ACCRA alcoholic beverage prices. In summary, the studies just mentioned provide a considerable amount of support for the utility of these prices in ongoing and future research.

Unfortunately, I cannot report that there are a large number of national micro panels to combine with the ACCRA prices. Of course, reduced-form versions of rational addiction models can be obtained by including past and future prices in cross-sectional demand functions. But these estimates do not take into account various restrictions implied by the models at issue. My research described in the fourth and fifth sections of this paper suggests that it is extremely important to incorporate these restrictions.

To my knowledge, there are only three U.S. national panels that include alcohol consumption. These are the panel component of "Monitoring the Future," the National Health and Nutrition Examination Survey I Followup (NHANESIF), and the National Longitudinal Survey of Youth (NLSY). Becker, Chaloupka, Murphy, and I are working with the first two data sets (Grossman et al. in progress). I will briefly describe these two databases and the work that we are doing with them. As yet, empirical results are not available. My aims are to indicate the kind of research that needs to be replicated and extended in other panels.

The NHANESIF, conducted between 1982 and 1984, was an attempt to reinterview the approximately 14,000 persons who underwent physical examinations in NHANESI and were between the ages of 25 and 74 at the time of these examinations (1971–75). Thus the age range of the roughly 11,000 respondents in the followup is between 32 and 86. Information on current alcohol consumption was collected in NHANESI, and information on current and lifetime consumption (in 10-year periods) was obtained in the followup.

With NHANESIF, two approaches to obtaining demand functions which incorporate rational addiction and employ a continuous measure of consumption will be explored. In the first approach, consumption reported in NHANESI, which was conducted approximately 10 years before NHANESIF, will be treated as current consumption; consumption during the past year in NHANESIF will be treated as future consumption (consumption 10 years later); and consumption 20 years ago reported retrospectively in NHANESIF will be treated as past consumption. In the second approach, past consumption and past price will be replaced as regressors by a measure of the addictive stock. For example, the addictive stock at age 55 of an NHANESIF respondent who is 65 years old will be computed based on his or her retrospective consumption measures at ages 25, 35, and 45 and on alternative assumed rates of depreciation.

Although we found that aggregate State data on per capita alcohol consumption is dominated by the behavior of moderate and light drinkers, this need not be the case in micro data. But we will obtain some demand functions after dividing the sample into heavy and nonheavy drinkers. This will be done to explore the prediction that the long-run price response should rise in absolute value as the degree of addiction rises.

Each year since 1975 the University of Michigan's Institute for Social Research has conducted a nationally representative random sample of between 15,000 and 19,000 high school seniors as part of the "Monitoring the Future" research program. These surveys focus on the use of illegal drugs, alcohol, and cigarettes. Starting with the class of 1976, a sample of 2,400 individuals in each senior class has been chosen for followup. Individuals in their senior year reporting current daily marijuana use or use of any other illegal drugs in the past 30 days are selected with a higher probability (by a factor of three). The 2,400 selected respondents are divided into 2 groups of 1,200 each; one group is surveyed on even-numbered calendar years, while the other group is surveyed on odd-numbered calendar years. Becker, Chaloupka, Murphy, and I are working with all waves through 1989 for the senior classes of 1976 through 1985. Thus our panels contain between three and seven waves of data inclusive of baseline for individuals aged 21 through 30 in 1989.[21]

We will obtain some demand functions for continuous measures of alcohol consumption with the Institute for Social Research panels. But in light of the rich nature of these panels, we will focus on applying insights from the rational addiction model to the decisions to start, restart, or quit heavy consumption using discrete-time hazard models. Clearly, start, restart, and quit probabilities are important outcomes to consider because of the addictive nature of the behavior at issue. Moreover, as indicated in the third section of this paper, the rational addiction model provides a theoretical frame-

[21] *The maximum number of waves is seven because there is no odd-year panel for the class of 1976.*

work to study start, quit, and restart decisions since unstable steady states are possible in this model. Thus, price changes can have large impacts on these probabilities.

First consider a model in which the rate of depreciation on the addictive stock is one. The start and restart probabilities in period t are negatively related to current price and positively related to consumption in period $t+1$ (or to the probability of participation in this period). These probabilities do not depend on consumption in period $t-1$ since it equals zero for persons who start or restart in period t. The quit probability in period t is positively related to current price and negatively related to consumption in periods $t-1$ and $t+1$ (or to the probabilities of participation in these two periods). Now consider a model in which the rate of depreciation is less than one. In such a model, future consumption cannot be used as a regressor. The reason is that the first-order conditions for utility maximization in three consecutive periods must hold to generate the structural demand function in this model. Not all these conditions can hold for persons who choose to quit or not to start or restart. The start probability is inversely related to current price and to future prices. The restart probability is inversely related to these variables and positively related to the current stock (which is positive for restarters since they engaged in the behavior before period $t-1$ and the depreciation rate is less than one). The quit probability falls as the current stock rises, and rises as current or future prices rise. Note that more than one future price may affect these probabilities because we have essentially solved for future consumption (replaced it by its determinants), and future consumption depends on more than one future price in the reduced form.

One difficulty in estimating the above models is that some of the right-hand side variables are endogenous, so that nonlinear simultaneous equations techniques are required. While we will experiment with these techniques, they may be very "taxing" on the data. Therefore, as an alternative, we will fit pure reduced form models in which past and future consumption are replaced by past and future prices in the model with a depreciation rate of one. Similarly, the stock is replaced by past prices in the model with a depreciation rate less than one.

In future research the NLSY could be used to replicate the results that Becker, Chaloupka, Murphy, and I obtain with the Institute for Social Research panels. The NLSY contains data on 12,000 youths who were between the ages of 14 and 21 in 1979. Data on alcohol use in this panel are available for the years 1984, 1987, and 1988.[22] While the institute panel is longer than the NLSY panel, the latter has much more detailed information on school and labor market outcomes. The institute panel itself remains a very promising source of data for future research as its members continue to age and new persons are added from high school senior classes after 1985. Persons 55 years and older in

[22] *Taubman et al. (1991) and Taubman and Sickles (1991) have developed research plans in which NLSY is used to estimate demand functions for legal and illegal drugs.*

NHANESIF were interviewed again in 1986, and all NHANESI survivors were interviewed in 1987. Alcohol consumption was obtained in each interview, and these two additional waves could be used to augment the research that Becker, Chaloupka, Murphy, and I are pursuing with NHANESIF. Other panels may be available in Canada and in countries in Western Europe. I can give no specific examples, but I suspect that a thorough search for panels may have substantial payoffs. Given the extremely interesting research with the National Health Interview Survey by Kenkel (in press), Manning et al. (1991), and Phelps and Leung (in progress), efforts to encourage the National Center for Health Statistics to add a panel component to this survey also may have substantial payoffs.

I want to mention two additional topics on an agenda for continuing and future research attention. One pertains to rational addiction studies of the effects on excessive alcohol consumption and cirrhosis mortality of increases in Federal excise tax rates on alcoholic beverages at the beginning of fiscal 1986 and in January 1991. An important message of the rational addiction model in this context is that the long-run impact of an excise tax hike will exceed the short-run impact. Research in this area could draw on micro panels, aggregate time series for the United States as a whole, or time series of State cross sections (provided time effects are not modeled with a full set of time dummies). The research with national data could modify interrupted time-series models, so that the outcome at issue depends on future events as well as on current and past events.

The second topic pertains to research that extends Showalter's (1991) model of the optimal pricing policy of monopolistic or oligopolistic producers of addictive goods to alcohol. This work would be particularly useful if it considered advertising as a decision variable by producers and obtained estimates of the effects of advertising on consumption in the context of rational and myopic models of addiction.

ACKNOWLEDGMENTS

Research for this paper was supported by grant 5 R01 AA08359 from the National Institute on Alcohol Abuse and Alcoholism to the National Bureau of Economic Research (NBER). This paper has not undergone the review accorded official NBER publications; in particular, it has not been submitted for approval by the board of directors.

I am grateful to Gregory Bloss, Frank Chaloupka, Michael Hilton, and Willard G. Manning for many helpful comments and suggestions.

REFERENCES

Babor, T.F.; Mendelson, J.H.; Greenberg, I.; and Kuehnle, J. Experimental analysis of the happy hour: Effects of purchase price on alcohol consumption. *Psychopharmacology* 58:35–41, 1978.

Baltagi, B.H., and Levin, D. Estimating dynamic demand for cigarettes using panel data: The effects of bootlegging, taxation and advertising reconsidered. *Rev Econ Stat* 68 (1):148–155, 1986.

Barthold, T.A., and Hochman, H.M. Addiction as Extreme-Seeking. *Econ Inquiry* 26(1):89–106, 1988.

Becker, G.S. "Optimal Discounting of the Future." Lecture, Department of Economics, University of Chicago, April 1990.

Becker, G.S. "Habits, Addictions, and Traditions." Nancy Schwartz Lecture, Northwestern University, June 1991.

Becker, G.S.; Grossman, M.; and Murphy, K.M. "An Empirical Analysis of Cigarette Addiction." National Bureau of Economic Research Working Paper No. 3322, Revised August 1991*a*.

Becker, G.S.; Grossman, M.; and Murphy, K.M. Rational addiction and the effect of price on consumption. *Am Econ Rev* 81(2):237–241, 1991*b*.

Becker, G.S., and Murphy, K.M. A theory of rational addiction. *J Political Econ* 96(4): 675–700, 1988.

Bigelow, G., and Liebson, I. Cost factors controlling alcoholic drinking. *Psychol Rec* 22:305–314, 1972.

Bover, O. Relaxing intertemporal separability: A rational habits model of labor supply estimated from panel data. *J Labor Econ* 9(1):85–100, 1991.

Boyer, M. A habit forming optimal growth model. *Int Econ Rev* 19(3):585–609, 1978.

Boyer, M. Rational demand and expenditures patterns under habit formation. *J Econ Theory* 31(1):27–53, 1983.

Chaloupka, F. "An Economic Analysis of Addictive Behavior: The Case of Cigarette Smoking." Ph.D. dissertation, City University of New York Graduate School, 1988.

Chaloupka, F. Rational addictive behavior and cigarette smoking. *J Political Econ* 99(4): 722–742, 1991.

Chaloupka, F.; Saffer, H.; and Grossman, M. Alcohol control policies and motor vehicle fatalities. *J Legal Stud*, in press.

Coate, D., and Grossman, M. Effects of alcoholic beverage prices and legal drinking ages on youth alcohol use. *J Law Econ* 31(1):145–171, 1988.

Cook, P.J., and Tauchen, G. The effect of liquor taxes on heavy drinking. *Bell J Econ* 13(2): 379–390, 1982.

Cook, P.J., and Tauchen, G. The effects of minimum drinking age legislation on youthful auto fatalities, 1970–1977. *J Legal Stud* 13(1):169–190, 1984.

Davidson, R., and McKinnon, J.G. Several tests for model specification in the presence of alternative hypotheses. *Econometrica* 49(3):781–793, 1981.

Deaton, A., and Muellbauer, J. *Economics and Consumer Behavior.* Cambridge: Cambridge University Press, 1980. 450 pp.

Donegan, N.H., et al. A learning theory approach to commonalities. In: Levinson, P.K.; Gerstein, D.R.; and Maloff, D.R.; eds. *Commonalities in Substance Abuse and Habitual Behavior.* Lexington, MA: D.C. Heath and Company, 1983. pp. 111–156.

Duffy, M. "Advertising, Taxation, and the Demand for Beer, Spirits, and Wine in the United Kingdom, 1963–1978." Occasional Paper No. 8009, Department of Management Sciences, UMIST, 1980.

El-Safty, A.E. Adaptive behavior, demand, and preferences. *J Econ Theory* 13:298–318, 1976.

Elster, J. *Ulysses and the Sirens: Studies in Rationality and Irrationality.* Cambridge: Cambridge University Press, 1979. 193 pp.

Farrell, P., and Fuchs, V.R. Schooling and health: The cigarette connection. *J Health Econ* 1(3):217–230, 1982.

Ghez, G.R., and Becker, G.S. *The Allocation of Time and Goods Over the Life Cycle.* New York: Columbia University Press for the National Bureau of Economic Research, 1975. 150 pp.

Grabowski, H.G. The effects of advertising on the interindustry distribution of demand. *Explor Econ Res* 3(1):21–75, 1976.

Grossman, M. Health benefits of increases in alcohol and cigarette taxes. *Br J Addict* 84(10):1193–1204, 1989.

Grossman, M.; Becker, G.S.; Chaloupka, F.; and Murphy, K.M. Alcohol, addiction, and price. National Institute on Alcohol Abuse and Alcoholism Grant No. 5 R01 AA08359, in progress.

Grossman, M.; Coate, D.; and Arluck, G.M. Price sensitivity of alcoholic beverages in the United States. In: Holder, H.D., ed. *Control Issues in Alcohol Abuse Prevention: Strategies for States and Communities* Greenwich, CT: JAI Press, Inc., 1987. pp. 169–198.

Hammond, P.J. Changing tastes and coherent dynamic choice. *Rev Econ Stud* 43(1):159–173, 1976.

Heckman, J. Life cycle consumption and labor supply: An explanation of the relationship between income and consumption over the life cycle. *Am Econ Rev* 64(1):188–194, 1974.

Heien, D., and Pompelli, G. Stress, ethnic and distribution factors in a dichotomous response model of alcohol abuse. *J Stud Alcohol* 48(5):450–455, 1987.

Hicks, J.R. *Capital and Growth.* Oxford: Oxford University Press, 1965. 339 pp.

Hotz, V.J.; Kydland, F.E.; and Sedlacek, G.L. Intertemporal preferences and labor supply. *Econometrica* 56(2):335–360, 1988.

Houthakker, H.S., and Taylor, L. *Consumer Demand in the United States: Analyses and Projections.* Second and Enlarged Edition. Cambridge, MA: Harvard University Press, 1970. 214 pp.

Iannaccone, L.R. "Consumption Capital and Habit Formation with an Application to Religious Participation." Ph.D. dissertation, University of Chicago, 1984.

Johnson, J.A., and Oksanen, E.H. Estimation of demand for alcoholic beverages in Canada from pooled time series and cross sections. *Rev Econ Statistics* 59(1):113–118, 1977.

Jones, A.M. A systems approach to the demand for alcohol and tobacco. *Bull Econ Res* 41(2):85–105, 1989.

Keeler, T.E.; Hu, T.; and Barnett, P.G. "Taxation, Regulation, and Addiction: A Demand Function for Cigarettes Based on Time-Series Evidence." Working Paper No. 91–173, Department of Economics, University of California at Berkeley, July 1991.

Kendell, R.E.; De Roumanie, M.; and Ritson, E.B. Influence of an increase in excise duty on alcohol consumption and its adverse effects. *Br Med J* 287:809–811, 1983.

Kenkel, D.S. Drinking, driving, and deterrence: The social costs of alternative policies. *J Law Econ*, in press.

Kennedy, K.A.; Ebrill, L.; and Walsh, B.M. The demand for beer and spirits in Ireland. *Proc Royal Irish Acad* 73(13), 1973.

Kydland, F.E., and Prescott, E.C. Time to build and aggregate fluctuations. *Econometrica* 50(6):1345–1370, 1982.

Lewit, E.M., and Coate, D. The potential for using excise taxes to reduce smoking. *J Health Econ* 1(2):121–145, 1982.

Lewit, E.M.; Coate, D.; and Grossman, M. The effects of government regulation on teenage smoking. *J Law Econ* 24(3):545–569, 1981.

MaCurdy, T.E. An empirical model of labor supply in a life-cycle setting. *J Political Econ* 89(6):1059–1085, 1981.

Manning, W.G.; Blumberg, L.; and Moulton, L.H. "The Demand for Alcohol: The Differential Response to Price." Working Paper, University of Minnesota, November 1991.

Marshall, A. *Principles of Economics.* Eighth Edition. London: The Macmillan Company, 1920. 871 pp.

Michaels, R.J. Addiction, compulsion, and the technology of consumption. *Econ Inquiry* 26(1):75–88, 1988.

Mobilia, P. "An Economic Analysis of Addictive Behavior: The Case of Gambling." Ph.D. disser-

tation, City University of New York Graduate School, 1990.

Nelson, J. State monopolies and alcoholic beverage consumption. *J Regul Econ* 2:83–98, 1990.

Pashardes, P. Myopic and forward looking behavior in a dynamic demand system. *Int Econ Rev* 27(2):387–397, 1986.

Peele, S. *The Meaning of Addiction: Compulsive Experience and Its Interpretation.* Lexington, MA: Lexington Books, 1985. 203 pp.

Pekurinen, M. The demand for tobacco products in Finland. *Br J Addict* 84(10):1183–1192, 1989.

Phelps, C.E., and Leung, S.F. Reducing drunk driving with taxes, laws, and information. National Institute on Alcohol Abuse and Alcoholism Grant No. 5 R01 AA08393, in progress.

Phlips, L. A dynamic version of the linear expenditure model. *Rev Econ Statistics* 54(4):450–458, 1972.

Phlips, L. The demand for leisure and money. *Econometrica* 46(5):1025–1043, 1978.

Phlips, L. *Applied Consumption Analysis.* Revised and Enlarged Edition. Amsterdam: North-Holland, 1983. 331 pp.

Pollak, R.A. Habit formation and dynamic demand functions. *J Polit Econ* 78(4/1):745–763, 1970.

Pollak, R.A. Habit formation and long-run utility functions. *J Econ Theory* 13:272–297, 1976.

Porter, R.H. The impact of government policy on the U.S. cigarette industry. In: Ippolito, P.M., and Scheffman, D.T., eds. *Empirical Approaches to Consumer Protection Economics.* Washington, DC: Federal Trade Commission, 1986. pp. 447–481.

Rimm, E.B.; Giovannucci, E.L.; Willett, W.C.; Colditz, G.A.; Ascheria, A.; Rosner, B.; and Stampfer, M.J. Prospective study of alcohol consumption and risk of coronary disease in men. *Lancet* 338:464–468, 1991.

Ryder, H.E., Jr., and Heal G.M. Optimum growth with intertemporally dependent preferences. *Rev Econ Stud* 40(121):1–33, 1973.

Saffer, H., and Grossman, M. Beer taxes, the legal drinking age, and youth motor vehicle fatalities. *J Legal Stud* 16(2):351–374, 1987*a*.

Saffer, H., and Grossman, M. Drinking age laws and highway mortality rates: Cause and effect. *Econ Inquiry* 25(3):403–418, 1987*b*.

Schelling, T.C. Egonomics, or the art of self-management. *Am Econ Rev* 68(2):291–294, 1978.

Schelling, T.C. *Choice and Consequence.* Cambridge, MA: Harvard University Press, 1984*a*.

Schelling, T.C. Self-command in practice, in policy, and in a theory of rational choice. *Am Econ Rev* 74(2):1–11, 1984*b*.

Scitovsky, T. *The Joyless Economy.* Oxford: Oxford University Press, 1976. 310 pp.

Showalter, M.H. "Monopoly Behavior with Intertemporal Demands." Ph.D. dissertation, Massachusetts Institute of Technology, 1991.

Spinnewyn, F. Rational habit formation. *Eur Econ Rev* 15(1):91–109, 1981.

Sumner, D.A. Measurement of monopoly behavior: An application to the cigarette industry. *J Polit Econ* 89(5):1010–1019, 1981.

Taubman, P.; Behrman, J.R.; and Pollak, R.A. Drugs and related demands: Prices and policies. Research Plan, National Bureau of Economic Research, January 1991.

Taubman, P., and Sickles, R. The demand for illegal drug use and prices. Research Plan, National Bureau of Economic Research, January 1991.

Thaler, R.H., and Shefrin, H.M. An economic theory of self-control. *J Polit Econ* 89(2):392–406, 1981.

Thom, R. The demand for alcohol in Ireland. *Econ Soc Rev* 15(4):325–336, 1984.

Townsend, J.L. Cigarette tax, economic welfare and social class patterns of smoking. *Appl Econ* 19:355–365, 1987.

Tsolakis, D.; Reithmuller, P.; and Watts, G. The demand for wine and beer. *Rev Marketing Agricultural Econ* 51(2):131–153, 1983.

von Weizsäcker, C.C. Notes on endogenous change of tastes. *J Econ Theory* 3:345–372, 1971.

Wasserman, J.; Manning, W.G.; Newhouse, J.P.; and Winkler, J. The effects of excise taxes and regulations on cigarette smoking. *J Health Econ* 10(1):43–64, 1991.

Winston, G.C. Addiction and backsliding. *J Econ Behav Organ* 1:295–324, 1980.

Young, T. The demand for cigarettes: Alternative specifications of Fujii's model. *Appl Econ* 15(2):203–211, 1983.

5

Advertising Under the Influence

Henry Saffer, Ph.D.

INTRODUCTION

Consumers are exposed to alcohol advertising from a number of sources. These sources include advertisements on broadcast television, cable television, and radio; in magazines and newspapers; on billboards; and through point-of-purchase promotions, consumer novelties, and sponsored events. In addition, there are also paid and unpaid portrayals of alcohol use in television dramas, films, literature, and popular music. The *Economist* (1990) reported that, in the United States, alcohol producers spend $1.3 billion a year on advertising and an equal amount on point-of-purchase promotions and sponsored events. Some public interest groups claim that these large expenditures on alcohol advertising increase alcohol consumption. The alcohol industry claims that alcohol advertising only affects market share.

Alcohol advertising has been the subject of a number of diverse studies by researchers from several disciplines. One type of study consists of experiments that examine how small groups, in controlled environments, react to controlled exposures to alcohol advertisements. Another type of study, which also relies on primary data collection, consists of in-depth interviews of small samples of individuals. In these interviews, data are collected on the media to which people have recently been exposed and on various measures of alcohol use. Advertising exposure data are then correlated with alcohol use data. Econometric studies rely on data from secondary sources, such as existing large-scale surveys of individuals and aggregate statistics for various communities. Econometric studies examine the effects of alcohol advertising on market share, industry profits, and alcohol consumption.[1] The results from a number of these types of studies are reviewed by Strickland (1983), van Iwaarden (1983), Atkin (1987), Smart (1988), Aitken (1989), and Waterson (1989). These reviews conclude that there is little evidence of any relationship between alcohol advertising and alcohol consumption.

The purpose of this chapter is to review econometric studies of alcohol

[1] *Content studies, another type of study, examine advertisements for consumer information, brand symbolism, and lifestyle portrayals. These studies usually do not correlate advertising exposure to consumer behavior.*

advertising effects on consumption and to provide some recommendations for future work in this area. This review has two novel aspects. First, it contains a discussion of the public health aspect of concentration in the alcohol industry. Second, it includes a review of econometric issues that are important in advertising research but which have not been employed extensively in prior alcohol advertising studies. In the next section, the role of advertising in the consumer demand curve is considered. The section that follows considers the effects of advertising on firms' market power and ability to set prices. Next, five important econometric issues in the estimation of the relationship between advertising and consumption are reviewed. A number of studies of alcohol advertising are reviewed with special attention given to how well these studies address the methodological issues considered earlier. The final section outlines some areas and possible methodological approaches for future work.

ADVERTISING AND THE CONSUMER'S DEMAND CURVE

The theoretical reason for including advertising in a consumer's demand curve is based on the informational aspect of advertising. Basmann (1956) argued that advertising can affect the functional form of the indifference curve. Schmalensee (1972) concluded that this argument violates the monotonic transformation assumptions of ordinal utility theory. Stigler and Becker (1977) have argued that advertising is an exogenous shift variable in the household production function, regardless of whether the information in the advertising is substantive or fanciful. In the Stigler-Becker model, advertising increases information and thus lowers the full price of the product. Advertising affects consumption by moving along an individual's demand curve rather than shifting the demand curve.

Even if the argument is accepted that consumer demand for a specific brand is affected by advertising, it does not necessarily mean that the consumer's demand for the industry product also is affected. Most advertising originates at the level of the individual firm and announces the virtues of specific brands rather than the industry product. Waterson (1989) pointed out that advertising managers have an incentive to use advertising that increases their firm's market share at the expense of the market share of other firms. Whether advertising affects the demand for a firm's product without influencing the industry demand, or affects both a firm's demand and the industry demand, is ultimately an empirical question.

ADVERTISING AND FIRM DEMAND

Many economic studies of advertising (e.g., see Schmalensee 1972 and Lambin 1976) are primarily concerned with the potential that advertising has to create monopoly power. The potential exercise of monopoly power by alcohol-producing firms is a public health issue if this power increases the industry price of alcohol. The hypothesis in this type of study is that advertising by a firm (referred to as firm A) shifts firm A's demand curve to the right at the expense of other firms. Firm

A's market share and concentration in the industry increase. Marginal revenue also shifts to the right, and firm A's equilibrium output and price rise. The stock of past advertising creates product visibility, or brand loyalty, that is expensive for rivals to overcome. The stock of past advertising also acts as a barrier to entry and would prevent new firms from reducing firm A's market share. However, some economists (e.g., see Boyd and Seldon 1990; Ashley et al. 1980) have concluded that the effects of advertising depreciate so rapidly that there is no accumulated stock of advertising. In this case, there is no advertising-created barrier to entry, and new firms could erode firm A's market share.

The extent to which firms can use advertising to gain monopoly power may have important public health implications. Successful advertisers can charge higher prices, which results in a lower industry-level quantity demanded. That is, advertising can raise product prices which lowers consumption and potentially reduces alcohol abuse. This effect would be offset to some degree if advertising induced increases in industry-level demand.

The empirical literature on the effects of advertising on individual firm demand is mixed. These empirical studies use concentration, profitability, and market share as outcome measures. Studies by Gisser (1991), Lustgarten and Thomadakis (1987), Mueller and Rogers (1980), and Grabowski (1978) indicated that firm advertising can increase monopoly power. Alternatively, studies by Eckard (1991, 1988) and Strickland and Weiss (1976) indicated that advertising does not result in monopoly power.

ADVERTISING AND INDUSTRY DEMAND

The public health aspect of alcohol advertising that has received the most attention is the effect of alcohol advertising on industry demand. Econometric evidence of an effect of advertising on industry demand, in any industry, is minimal. However, some researchers (e.g., Mosher and Wallack 1981) contend that econometric research is flawed and should be discounted. They argue that econometric studies measure only small marginal effects of alcohol advertising on industry demand. Observations on advertising expenditures that cover a wider range would provide a more accurate estimate of the effect of advertising.

ECONOMETRIC ISSUES

Before reviewing the econometric results from studies of alcohol advertising and industry demand, there are five econometric issues that are worth noting. [See Berndt (1991, chapter 8) for a more extensive review of these issues.]

The first econometric issue involves specification of the relationship between advertising and consumption. Although there have been a number of theoretical and empirical studies of advertising and brand sales, there is no consensus on the functional relationship. A study by Ackoff and Emshoff (1975) provides some insight into the relationship between brand advertising and consumption. They hypothesized that beer advertising is sub-

ject to diminishing returns. Their experiments conducted with the Anheuser-Busch Company provided support for this theory. Another brand-level empirical study by Lambin (1976) of several consumer nondurable goods also supports this theory. Alternatively, Dominique (1988) provided a model which concludes that there are increasing returns to advertising. An empirical study by Peles (1971) also demonstrated increasing returns to beer advertising.

These studies suggest that if there is any relationship between aggregate advertising and industry sales, it may be a nonlinear one: In econometric models, quadratic or logarithmic specifications should be examined. If the relationship is quadratic, then omission of the squared advertising term would bias the coefficient of the linear term toward zero. If the relationship is logarithmic, estimation of a linear model will result in specification bias in the coefficients of the included variables.

The second econometric issue is simultaneity between advertising and consumption. Some economists (see Schmalensee 1972) argue that, for mature products, advertising budgets are adjusted so that advertising expenditure is a fixed percentage of consumption. On the other hand, firms undertake advertising so that they can increase consumption. These assumptions imply that advertising and consumption are jointly determined. If this is the case, ordinary least squares (OLS) estimation will be biased, and a two-stage estimation model is needed. A two-stage estimation model must have a structural consumption equation and structural advertising equation. The structural advertising equation is a demand for advertising equation. Each structural equation must be identified by at least one exogenous variable that does not appear in the other equation. Identification of the consumption equation is achieved relatively easily with product price and other variables. The advertising structural equation is harder to identify. Although advertising price and consumption response measures are candidates, these data are difficult to collect. Finally, in models that treat advertising as endogenous, it is empirically unrealistic to attempt to use more than one advertising variable.

If advertising is thought to be endogenous, a test for endogeneity is also important. The Wu-Hausman test for simultaneity can be performed by first estimating the reduced form advertising equation. The predicted value of the dependent variable is added to the structural consumption equation, which is then estimated by OLS. If the coefficient of the predicted advertising variable is significant, then advertising cannot be considered exogenous. This test presumes an identified structural model.

The third econometric issue in estimating the effects of alcohol advertising on alcohol consumption involves controlling for the possible cumulative effects of advertising. This possibility suggests a model in which advertising is treated as a stock. The stock of advertising decays over time and increases with new exposures. In this model it is assumed that a

series of lagged advertising variables affect current consumption. One way to estimate this model is to assume that the relationship between the coefficients of the lagged advertising variables can be represented as $\beta_{t+1} = \lambda\beta_t$, with $0 > \lambda \geq 1$. Using a Koyck transformation, an estimable version of this model emerges. The empirical equation includes current advertising and lagged consumption as independent variables. The coefficient of the advertising variable, β_0, measures the contemporaneous effect of advertising on consumption. The coefficient of the lagged consumption variable, λ, measures the carryover effect of the prior period's advertising. This model must be estimated by maximum likelihood or instrumental variable methods due to correlation of the error term with lagged consumption.

This model can be used to estimate the cumulative and lingering effects of advertising. The cumulative effects of advertising are estimated as $\beta_0(1-\lambda^m)/(1-\lambda)$, where m is the number of periods of advertising. The lingering effect of advertising can be defined as the proportion of the total consumption response that persists after m periods. In this model, $p = (1-\lambda^m)$, where p is the proportion of total consumption response that persists m periods after the advertising.

The fourth econometric issue in estimating the relationship between alcohol consumption and advertising is temporal aggregation bias. Temporal aggregation bias refers to bias resulting from the length of the time interval used to measure the data. Aggregation over a long interval may obscure important variance in the dependent and independent variables. Alcohol advertising expenditures probably should be measured in monthly or quarterly intervals. These expenditures follow a seasonal pattern and also vary during the year due to expenditure pulsing schedules. Expenditure pulsing is a technique of spending a given annual budget in short, intense intervals of advertising rather than maintaining a constant level of spending throughout the year. The effects of these expenditure pulses, which may linger for a month or more, result in maximizing the effect of the given budget. In both the case of seasonal spending and of pulsing, annual data will not measure these important variations in expenditure and any resulting influence on consumption.

Maddala (1977) discussed temporal aggregation bias in the context of distributed lag models. Also, a study by Bass and Leone (1983) showed that λ decreases and β_0 increases with the length of the time interval used to measure the data. A Monte Carlo study by Weiss and associates (1983) found that a data interval of less than a year is necessary for studies of the effects of advertising of nondurable consumer goods.

The last econometric issue involves the measurement of advertising. In estimating a demand curve using aggregate data, the advertising variable should measure the number of effective advertising messages in the aggregation group. Effectiveness refers to the number of people exposed to an advertising message and the impact of the message. The advertising data that are generally available are

measures of advertising expenditures. Real expenditures can be estimated with the consumer price index or the McCann-Erickson advertising price index. If it is assumed that the price of advertising is proportional to effectiveness, then real expenditure data are proportional to effective messages. Real expenditures for advertising in different media cannot be added together unless it is also assumed that the relationship between effectiveness and price is the same for all the various advertising media.

To summarize, five econometric issues important in advertising research are: (1) testing of nonlinear specifications, (2) simultaneous equations bias, (3) cumulative effects of advertising, (4) temporal aggregation bias, and (5) empirical proxies for effective advertising messages. The possibility of implementing or testing these issues varies with the type of data employed. Testing nonlinear specifications is relatively straightforward and should be possible with all data sets. Testing for simultaneous equations bias depends on being able to identify an advertising demand curve; however, variables that could be used for this purpose are not always available. The next two issues of cumulative effects and temporal aggregation bias are related: If the impact of an advertising message decays quickly, then the data must be collected over short time intervals. Finally, not many proxies for effective advertising messages are available in most data sets. Although advertising expenditure data are a reasonable alternative, identifying new proxies for advertising messages is an important goal for future studies.

PRIOR STUDIES OF ALCOHOL ADVERTISING AND INDUSTRY DEMAND

Over the past few years, a number of econometric studies have been conducted on the effect of alcohol advertising on industry-level demand. These studies have, on the whole, provided little evidence that alcohol advertising increases the industry-level demand for alcohol. One type of study relies on a short time series of annual data from one country. Advertising is measured by expenditures, and control variables such as price and income are generally included. The studies by Grabowski (1976), McGuiness (1980), Duffy (1987), and Selvanathan (1989) are representative of this group.[2]

Grabowski (1976) studied the effect of advertising on consumer expenditure on alcohol. Seventeen annual observations were taken over the time period 1956–72. He tested logarithmic specifications, used a simultaneous model to account for endogeneity, and included lagged expenditure to account for the stock aspect of advertising. Grabowski thus accounted for nonlinearities, simultaneity, and cumulative effects. Use of annual data is a problem in this work, as is the very small sample size. He found that advertising has no effect on alcohol expenditures but that expenditure has a positive effect on alcohol advertising.

McGuiness (1980) examined the effect of print and television advertising on alcohol consumption. The time period

[2] *See also Comanor and Wilson (1974), Walsh (1982), McGuiness (1983), Duffy (1983), and Johnson (1985).*

included 20 annual observations from 1956 to 1975 for the United Kingdom. Separate measures for beer, wine, and spirits advertising were used. Logarithmic models were tested, and a simultaneous model of consumption and advertising was estimated. McGuiness thus accounted for nonlinearities and simultaneity. Annual data and cumulative effects are problems as is the very small sample size. He found that spirits advertising has an effect on total alcohol consumption.

Duffy (1987) used a Rotterdam model specification to examine the effect of television and print advertising on alcohol expenditures. The data included 21 annual observations from 1963 to 1983 for the United Kingdom. The Rotterdam specification is nonlinear but imposes additional constraints on the estimated parameters. Simultaneity, cumulative effects, and temporal aggregation bias are problems. He found that beer advertising has an effect on alcohol expenditures.

Selvanathan (1989) used a systemwide analysis of the effects of advertising on the demand for beer, wine, and spirits in the United Kingdom for the period 1955–75. He accounted for nonlinearities, simultaneity, and cumulative effects. The results showed that beer advertising increases beer consumption and decreases wine and spirits consumption. Moreover, both wine and spirits advertising decrease the consumption of beer. The analysis is extended by considering advertising as a stock rather than a flow. The findings indicated that the advertising of alcoholic beverages fully depreciates in consumers' minds in 1 year. These results are somewhat surprising since annual data are employed in the estimation models. Again, the very small sample size is a problem.

These four studies represent important efforts in estimating the effects of alcohol advertising. The results are, however, weak and inconsistent. The limitations of a short annual time series for a single country leave some doubts about the utility of pursuing this type of data in future studies.

Three additional studies improve somewhat upon the data problems. A study by Duffy (1991) and a study by Franke and Wilcox (1987) used quarterly data. Another study by Bourgeois and Barnes (1979) added some novelty by using cross-sectional data. The results in all three studies, however, do not provide support for the hypothesis that advertising increases industry-level demand.

Duffy (1991) extended the Rotterdam type of empirical model to include four aggregate categories of consumer nondurables, one of which was alcoholic drink and tobacco. Duffy argued that better estimates of the effect of advertising on the interproduct distribution of demand are found with a set of product demand curves. The data are quarterly observations for the United Kingdom for the period 1963–85. Nonlinearities and temporal aggregation bias were addressed, although simultaneity and cumulative effects were not considered. He concluded that advertising has no effect on industry-level demand.

Franke and Wilcox (1987) examined the relationship between industrywide

advertising expenditures and per capita consumption of beer, wine, and distilled spirits in the United States. Advertising expenditures were analyzed for six different media: network television, spot television, network radio, magazines, newspaper supplements, and billboards. Nonlinearities, simultaneity, and cumulative effects were not considered. The results showed no evidence of a significant relationship between total advertising and consumption of beer. However, positive, significant relationships were found between advertising and the consumption of wine and distilled spirits.

Bourgeois and Barnes (1979) used data for 10 Canadian provinces for the period 1951–74. However, advertising data were available for only 5 of the years in the data set. In 4 of these 5 years the data were limited to five regions. This apparently leaves only 30 observations. Nonlinearity, simultaneity, and cumulative effects were not considered. Temporal aggregation bias is also a problem because annual data were used. The authors found that neither print nor broadcast advertising has any effect on total alcohol consumption.

Another type of study by Saffer (1993) used market-level data to test the effects of alcohol advertising on motor vehicle fatality rates. The unit of observation in this data set is an area of dominant influence (ADI), as defined by the Arbitron Company. The ADI, a well-known concept in advertising research, consists of an aggregation of counties within the reach of a given group of television transmitters. The original ADI definitions were established in 1965 and are updated annually. The ADI is similar to, but somewhat larger than, a metropolitan statistical area. The data consisted of quarterly observations from 1986 to 1989 for the top 75 ADI's in the United States. These 75 television markets account for more than 75 percent of the total U.S. population. The dependent variables in this study consisted of a series of motor vehicle fatality rates computed with data from the Fatal Accident Reporting System. A logistic transformation of the dependent variables was used. The independent variables consisted of alcohol advertising expenditures on television, radio, and billboards. Other independent variables on the ADI level included alcohol prices, personal income, and demographic information. Regional dummies were also included in the regressions. Results from regression models that account for endogeneity of advertising expenditures showed that advertising increases motor vehicle fatalities.

Another group of studies (Smart and Cutler 1976; Ogborne and Smart 1980; Ornstein and Hanssens 1985; Seldon and Doroodian 1989; Saffer 1991) used advertising bans as the measure of advertising. To measure the effect of a ban, a control group is needed. Depending on the data set, the control can be a geographic area without a ban, or the control can be a time period without a ban. In the case of an alternative aggregate, additional appropriate controls for other factors are necessary. In the case of a time series, the data should cover a long time period in order to capture any adjustments to a ban. The effects of advertising bans have been stud-

ied with interrupted time series techniques and in regression models.

Measuring the effect of a ban is a useful empirical approach for three reasons. First, alcohol advertising bans are used in a number of countries as a part of alcohol control policy. Second, a ban represents a larger variation in advertising exposures than does an expenditure variable. One criticism of econometric research using advertising expenditures to measure advertising is that only the marginal effects of small changes in advertising expenditures are estimated. If, in the empirically observed range, advertising expenditures are subject to diminishing returns, then a regression coefficient measuring the effect of advertising on consumption will underestimate the average effect. An advertising ban does not have this problem because it is an empirical proxy for expenditures measured over a large range. Third, an alcohol advertising ban may also be interpreted as an information signal. An alcohol advertising ban signals society's collective disapproval of alcohol abuse. This aspect of a ban could have a negative effect on alcohol consumption even if alcohol advertising had no effect on total alcohol consumption.

The econometric issues associated with measuring the effects of an advertising ban are somewhat similar to the problems associated with expenditure or message variables. Nonlinearity in the dependent variable should be tested. The major issue is whether the ban was enacted as a response to concern over alcohol abuse. This possibility can be tested with a Wu-Hausman test for endogeneity, which presumes an identified structural model. Since the ban is measured by a dummy variable, the ban equation should be estimated by using a probit model. The alcohol consumption equation would include an endogenous dummy advertising ban variable, and the *t* values must be appropriately adjusted. The issue of cumulative effects should be addressed by measuring the effects of a ban a number of time periods after enactment. Temporal aggregation bias is not an issue because the effect of a ban should be measured over a long time period.

Studies of alcohol advertising bans using interrupted time-series techniques were conducted by Smart and Cutler (1976) and Ogborne and Smart (1980). These studies examined the effects of advertising bans in British Columbia and Manitoba, respectively. The ban in British Columbia included all alcohol advertising but lasted only 1 year. The ban in Manitoba included only beer advertising and was analyzed using an 8-year time series. Both studies concluded that these advertising bans had no effect on alcohol consumption. Alternatively, these results may indicate that in a study of a single province or country, a long time period is required before changes in alcohol consumption are observable.[3]

In another study of advertising bans, Ornstein and Hanssens (1985) examined

[3] *The provincial bans may not have resulted in much of a reduction in total advertising exposure since these provinces receive a considerable amount of television programming from the United States. I owe this point to Dr. Louis Glicksman of the University of Western Ontario.*

the effect of laws banning billboard advertising, consumer novelties, and price advertising on the consumption of beer and spirits in the United States. State data for the period 1974–78 were used. The results showed that States that allow price advertising and consumer novelties have higher spirits consumption. They also found that billboard advertising and novelties have no effect on beer consumption, whereas there is some evidence that price advertising increases beer consumption.

Saffer (1991) examined the effect of banning broadcast advertising of alcoholic beverages on alcohol abuse. This study contains the first set of estimates of the effect of television advertising bans on alcohol abuse based on international data. The effect of a ban is difficult to estimate using data from one country because the adoption of new advertising bans is an infrequent event and requires many years for adjustment. However, an international data set avoids this problem because there is considerable variation in the use of advertising bans across countries. The data used in this study were a pooled time series from 17 countries for the period 1970–83. The empirical measures of alcohol abuse were per capita alcohol consumption, liver cirrhosis fatality rates, and highway fatality rates. The many cultural factors which influence alcohol use were measured by alcohol production variables and a set of country dummy variables. The empirical results showed that both alcohol advertising bans and alcohol price can have significant effects in reducing alcohol abuse.

A related study by Seldon and Doroodian (1989) of the advertising ban on cigarettes in the United States also found that an advertising ban reduces aggregate consumption. The authors used a simultaneous cigarette demand function and advertising function and estimated the model using time-series data and nonlinear three-stage least squares. This study also found that the increased health warnings about cigarettes resulted in increased industry advertising.

FUTURE ECONOMETRIC STUDIES OF ALCOHOL ADVERTISING

An area of study which has received little attention is the effect of advertising on concentration in the beer, wine, and spirits industries. Over time, advertising may have led to increased concentration and higher prices in these industries. To study this issue, a time-series data set with advertising, concentration, profitability, and price data would have to be constructed. Firm-level data would be useful because concentration will result from differential levels of advertising by firms rather than total industry advertising.

Several public interest groups have criticized alcohol producers for targeting advertising at blacks, Hispanics, and young people. Targeting, a common method of optimizing an advertising budget, refers to both the advertising content and the choice of media used to convey the advertising. Public interest groups contend that alcohol advertising increases drinking and thus exacerbates social problems in the targeted groups. Alcohol producers contend that their advertising has

no causal influence on the level of alcohol abuse. They also contend that there is no targeting of individuals under the legal drinking age since the industry has adopted voluntary advertising guidelines which prohibit such advertising. Targeted advertising might be a public health concern if this advertising is measurably larger in some target groups and can be causally linked to measurably larger public health problems in these target groups.

Measuring targeted advertising with available data may be difficult. However, with many of the data sets described above it is possible to estimate a demand for advertising equation. Age, race, and sex data, if available by unit of observation, can be included in the advertising demand equation. The other independent variables in this equation would control for other factors influencing the level of advertising in each unit of observation. Significant and positive coefficients on the demographic variables would provide some evidence of targeting.

Another approach to targeting would be to study selected media such as magazines, billboards, non-English language newspapers, and broadcasting. Because magazines are targeted to specific audiences, magazine data offer the best possibility of measuring targeted advertising. Alcohol advertising by specific magazine, by month, and by region are available. Annual demographic profiles for all major magazines are also available.[4] A data set could be constructed, using the magazine as the unit of observation, that includes alcohol advertising expenditures and demographic characteristics. This data set could be used to test for disproportionate advertising expenditures by demographic category. However, if it were found that any groups were disproportionately exposed, it would still be necessary to show that there is a causal link between alcohol advertising and alcohol abuse.

Future studies of the effect of alcohol advertising on industry demand might try to exploit cross-sectional data and pooled cross-sectional time-series data more fully. One type of potentially interesting cross-sectional data is micro data derived from large-scale surveys of individuals. A number of these surveys include measures of alcohol consumption, alcohol abuse, general health, income, employment, and demographics. Prior studies have not resolved the question of whether alcohol advertising affects demand. Future studies might be more definitive if more extensive use is made of alternative alcohol abuse measures. Another type of cross-sectional data is market-level data constructed around the reach of local media such as a broadcast television market. This type of data has been used by Saffer (1993). Three other types of cross-sectional data are county level, State level, and international data. Examples of the use of these are Ornstein and Hanssens (1985), who used a State-level data set, and Saffer (1991), who used an international data set.

[4] *The magazine data are available from the Publishers Information Bureau of Leading National Advertisers. The demographic data on magazine readership are available from Simmons Market Research Bureau. See Warner et. al. (1992) for an interesting application of these data to cigarette advertising.*

One reason that these types of data sets have not been used more extensively is the difficulty of measuring advertising on the cross-sectional level. Data on aggregate annual advertising expenditures are readily available from published sources. However, advertising industry sources are needed to use micro data sets or market-level data sets. Several commercial firms regularly monitor all types of alcohol advertising, and these data are available to researchers. These data can be used to construct market-level advertising variables by beverage type for television, radio, billboards, and newspapers. Alcohol advertising in magazines is also available on the regional level.

Micro data sets that contain data on alcohol consumption can be used to study the effects of alcohol advertising if the respondent's county of residence can be identified. When the respondent's county of residence can be identified, market-level advertising can be appended to the individual data. One advantage of a micro data set is that simultaneity between advertising and consumption is not an issue. Some micro data sets also have information on lagged consumption, which allows for estimation of cumulative effects. Other variables sometimes found in micro data sets are measures of drunk driving and information on media exposure. Generally, an important missing variable in this type of data set is media exposure. If media exposure data are not available, demographic data can act as proxies to some extent. Micro data are particularly useful in the study of alcohol consumption and alcohol abuse by age and other demographic characteristics.

A few micro data sets have followed a panel of individuals over time (e.g., the National Longitudinal Study, the National Health Interview Survey, and the National Health and Nutrition Survey). These panels have been reinterviewed on a regular basis for many years. This type of data has the same advantages as a single cross section, as well as the advantage of adding a time-series dimension. The time-series aspect allows for construction of first-difference models which eliminate time-invariant personal characteristics. A first-difference model would also eliminate the time-invariant error in self-reported alcohol consumption. In addition, alternative measures of cumulative advertising exposure can be constructed. Because some individuals are lost over time, sample selection must be controlled.

Market-level data sets also can be used to evaluate the effects of alcohol advertising on alcohol consumption and abuse. In a market-level data set, the unit of observation is the geographic area reached by an influential medium such as broadcast television. The geographic area for each market can then be defined as a specific group of counties and thus approximates a metropolitan area. This type of data can also be constructed over time, which considerably increases the size of the data set. The time-series aspect of market-level data also allows for computation of cumulative-effects models with a variety of cumulative advertising exposure variables.

In a market-level data set, endogeneity of both consumption and advertising

becomes an issue. Endogeneity can be tested for, and corrected for, with a simultaneous model. This requires estimation and identification of an advertising demand curve. The advertising demand curve can be identified with variables that measure the price of advertising or market-level variations in sales potential or advertising response.

Market-level data sets may also be useful in measuring the effects of public service announcements. Advertising trade associations regularly sample the top 75 television markets and can provide data on the number and extent of antialcohol advertisements. Because this type of advertising is generally of local origin, there is likely to be considerable variation in the data. A public service announcement variable could be added as an independent variable in a regression with highway fatalities or alcohol consumption as dependent variables.

The dependent variables appropriate for use with market-level data sets include alcohol consumption as well as mortality from cirrhosis, alcohol dependence, nondependent abuse, alcohol psychosis, alcohol poisoning, motor vehicle accidents, alcohol-related suicide, and alcohol-related homicide. In market-level data sets, independent variables should be measured at the county and the State levels. County-level data can be aggregated using population weights to create market-level variables. Some State-level data, such as alcohol excise taxes and State laws, have the same values for all locations in the State. If the State data are of this type, they can be used in the market-level data set. For markets that cross State boundaries, population-weighted data can be computed.

County-level data also may be useful in evaluating the effects of advertising on alcohol abuse. This type of data set should be limited to counties for which there are broadcast television data available. Only about half the counties, but over 75 percent of the U.S. population, would be included in such a data set. The advertising and price data used with a county-level data set would have very little additional variance over a market-level data set. The advantage of a county-level data set is increased variance in the dependent variables and demographic variables. With the exclusion of alcohol consumption, the variables available on the market level are also available on the county level. State-level variables, such as State laws and alcohol excise taxes, can also be included in a county-level data set. To create a county-level data set, some assumptions would have to be made about some forms of advertising data, such as billboard data, which are reported on the city or town level.

State-level data sets can be used to evaluate the effects of State laws regulating alcohol advertising. There is a wide variety in the mix of alcohol advertising regulatory options adopted in different States. These regulations include laws that restrict or limit billboard and newspaper advertising, point-of-sale advertising, and consumer novelties. These laws also differ by beverage category; typically, the most restrictive laws apply to spirits. The sample size and variance in State-level data

sets can be increased by pooling several years of data. The dependent variables that are available include alcohol consumption and a variety of alcohol-related mortality rates. State-level alcohol taxes, income, and demographics are also easily available. Endogeneity of State laws is a possibility that should be examined. For example, the study by Ornstein and Hanssens (1985) could be extended with more advertising data, more outcome variables, and more time periods of data.

International data sets can be constructed to study the effects of advertising bans. A number of countries have adopted alcohol advertising bans, which differ by media type and beverage type. Advertising bans are important to examine since they are a likely policy choice for regulation of alcohol advertising. Several alcohol outcome measures, such as alcohol consumption, liver cirrhosis mortality rates, and highway mortality rates, are available as dependent variables. Various price, income, and demographic data are also available by country. Again, simultaneity between alcohol use and advertising bans should be examined. The study by Saffer (1991) could be extended with more independent variables and with a data set that covered more countries and included more recent data.

In summary, new econometric studies could examine concentration and targeting and use alternative data sets to estimate the effects of advertising on alcohol abuse. Concentration studies could examine the question of whether advertising has increased concentration, raised prices, and through higher prices decreased consumption. Targeting studies could examine the question of whether youths and minorities are differentially exposed to advertising and whether these demographic groups are differentially influenced by alcohol advertising. Alternative data sets which could be used to estimate the effects of advertising on various alcohol abuse measures include microdata sets, market-level data sets, county-level data sets, State-level data sets, and international data sets.

REFERENCES

Ackoff, R., and Emshoff, J. Advertising research at Anheuser-Busch, Inc. (1963–68). *Sloan Manage Rev* 16(2):1–15, 1975.

Aitken, P. Television alcohol commercials and under-age drinking. *Int J Advertising* 8:133–150, 1989.

Ashley, R.; Granger, C.; and Schmalensee, R. Advertising and aggregate consumption: An analysis of causality. *Econometrica* 48:1149–1168, 1980.

Atkin, C. Alcoholic-beverage advertising: Its content and impact. In: Holder, H., ed. *Control Issues in Alcohol Abuse Prevention: Strategies for States and Communities.* Greenwich, CT: JAI Press Inc., 1987. pp. 267–287.

Basmann, R. A theory of demand with variable consumer preferences. *Econometrica* 24:47–58, 1956.

Bass, F., and Leone, R. Temporal aggregation, the data interval bias, and empirical estimation of bimonthly relations from annual data. *Manage Sci* 29:1–11, 1983.

Berndt, E. *The Practice of Econometrics: Classic and Contemporary.* Reading, MA: Addison-Wesley Publishing Co., 1991. 702 pp.

Bourgeois, J., and Barnes, J. Does advertising increase alcohol consumption? *J Advertising Res* 19(4):19–29, 1979.

Boyd, R., and Seldon, B. The fleeting effect of advertising. *Econ Lett* 34:375–379, 1990.

Comanor, W., and Wilson, T. *Advertising and Market Power.* Cambridge: Harvard University Press, 1974. 257 pp.

Dominique, C. A note on increasing returns to advertising. *Econ Lett* 28:381–385, 1988.

Duffy, M. The demand for alcoholic drink in the United Kingdom, 1963–78. *Appl Econ* 15(1):125–140, 1983.

Duffy, M. Advertising and the inter-product distribution of demand. *Eur Econ Rev* 31:1051–1070, 1987.

Duffy, M. Advertising in demand systems: Testing a Galbraithian hypothesis. *Appl Econ* 23:485–496, 1991.

Eckard, E. Advertising, concentration changes and consumer welfare. *Rev Econ Stat* 2:340–343, 1988.

Eckard, E. Competition and the cigarette TV advertising ban. *Econ Inquiry* 29:119–133, 1991.

Economist Magazine. Smoking 'em out. 316(7672):83, 1990.

Franke, G., and Wilcox, G. Alcoholic beverage advertising and consumption in the United States, 1964–1984. *J Advertising* 16(3):22–30, 1987.

Gisser M. Advertising, concentration and profitability in manufacturing. *Econ Inquiry* 29:148–165, 1991.

Grabowski, H. The effects of advertising on the interindustry distribution of demand. *Explorations Econ Res* 3(1):21–75, 1976.

Grabowski, H. The effects of advertising on the intraindustry shifts in demand. *Explorations Econ Res* 4(5):675–701, 1978.

Johnson, L. Alternative econometric estimates of the effects of advertising on the demand for alcoholic beverages in the United Kingdom. *Int J Advertising* 4(1):19–25, 1985.

Lambin, J. *Advertising, Competition and Market Conduct in Oligopoly Over Time.* Amsterdam: North-Holland, 1976. 312 pp.

Lustgarten, S., and Thomadakis, S. Mobility barriers and Tobin's q. *J Business* 60(4):519–537, 1987.

Maddala, G. *Econometrics.* New York: McGraw-Hill, 1977. 516 pp.

McGuinness, T. An econometric analysis of total demand for alcoholic beverages in the U.K. 1956–1975. *J Ind Econ* 29(1):85–105, 1980.

McGuinness, T. The demand for beer, spirits and wine in the UK, 1956–1979. In: Grant, M.; Plant, M.; and Williams, A., eds. *Economics and Alcohol.* New York: Gardner Press, Inc., 1983. pp. 238–242.

Mosher, J., and Wallack, L. Government regulation of alcohol advertising: Protecting industry profits versus promoting the public health. *J Public Health Policy* 2(1):333–353, 1981.

Mueller, W., and Rogers, R. The role of advertising in changing concentration of manufacturing industries. *Rev Econ Stat* 62(1):89–96, 1980.

Ogborne, A., and Smart, R. Will restrictions on alcohol advertising reduce alcohol consumption? *Br J Addict* 75:293–296, 1980.

Ornstein, S., and Hanssens, D. Alcohol control laws and the consumption of distilled spirits and beer. *J Consumer Res* 12:200–213, 1985.

Peles, Y. Economies of scale in advertising beer and cigarettes. *J Business* 44(1):32–37, 1971.

Saffer, H. Alcohol advertising bans and alcohol abuse: An international perspective. *J Health Econ* 10:65–79, 1991.

Saffer, H. "Alcohol advertising and motor vehicle fatalities." Unpublished manuscript,

National Bureau of Economic Research, New York, 1993.

Schmalensee, R. *The Economics of Advertising.* Amsterdam: North-Holland, 1972. 312pp.

Seldon, B., and Doroodian, K. A simultaneous model of cigarette advertising: Effects on demand and industry response to public policy. *Rev Econ Stat* 71(4):673–677, 1989.

Selvanathan, E. Advertising and alcohol demand in the UK: Further results. *Int J Advertising* 8:181–188, 1989.

Smart, R. Does alcohol advertising affect overall consumption? A review of empirical studies. *J Stud Alcohol* 49(4):314–323, 1988.

Smart, R., and Cutler R. The alcohol advertising ban in British Columbia: Problems and effects on beverage consumption. *Br J Addict* 71:13–21, 1976.

Stigler, G., and Becker, G. De gustibus non est disputandum. *Am Econ Rev* 67:76–90, 1977.

Strickland, D. Advertising exposure, alcohol consumption, and misuse of alcohol. In: Grant, M.; Plant, M.; and Williams, A., eds. *Economics and Alcohol.* New York: Gardner Press, 1983. pp. 201–222.

Strickland, A., and Weiss, L. Advertising, concentration, and price-cost margins. *J Polit Econ* 84(5):1109–1121, 1976.

van Iwaarden, M. Advertising, alcohol consumption, and policy alternatives. In: Grant, M.; Plant, M.; and Williams, A., eds. *Economics and Alcohol.* New York: Gardner Press, 1983, pp. 223–237.

Walsh, B. The demand for alcohol in the UK: A comment. *J Ind Econ* 30(4):439–446, 1982.

Warner, K.; Goldenhar, L.M.; and McLaughlin, C.G. Cigarette advertising and magazine coverage of the hazards of smoking—A statistical analysis. *N Engl J Med* 326(5):305–309, 1992.

Waterson, M.J. Advertising and alcohol: An analysis of the evidence relating to two major aspects of the debate. *Int J Advertising* 8:111–131, 1989.

Weiss, D.; Weinberg, C.; and Windal, P. The effects of serial correlation and data aggregation on advertising measurement. *J Marketing Res* 20:91–101, 1983.

6

Alcohol and the Labor Market

John Mullahy, Ph.D.

. . . the vast majority of modern drinkers do not feel any the worse for it, because they do not miss the extra efficiency they would enjoy on the water waggon. Very few people are obliged by their occupations to work up to the extreme limit of their powers. Who cares whether a lady gardener or a bookkeeper or a typist or a shop assistant is a teetotaller or not, provided she always stops short of being noticeably drunk? It is to the motorist or the aeroplane pilot that a single glass of any intoxicant may make the difference between life and death. . . . What would be sobriety for a billiard marker would be ruinous drunkenness for a professional billiard player.

George Bernard Shaw
The Intelligent Woman's Guide to Socialism, Capitalism, Sovietism and Fascism, p. 397

INTRODUCTION

Concerns about the effects of drinking[1] on social welfare are centuries old (for excellent historical overviews, see Aaron and Musto 1981 and Heather and Robertson 1989). Earlier attention to the morality of drinking has yielded to modern preoccupation with vehicular accidents related to drunk driving, youth drinking, fetal alcohol syndrome, cost-effective treatment of alcoholism, and—the topic of this chapter—the effects of drinking on success in the labor market. Yet, while outright Prohibition has given way to the complex web of regulations and taxes that form the substance of this generation's public policies on alcohol use, the fundamental premise of policy deliberations endures: Preventing problems that attend the misuse of alcoholic beverages is preferable to curing them once they have transpired.

Prevention, of course, is not without cost.[2] Accordingly, it is useful to have

[1] *Unless otherwise noted, "drinking" means consumption of alcoholic beverages.*
[2] *See Russell (1986) for a lucid discussion of the "prevention vs. cure" debate. While it has historically been presumed that an ounce of prevention is worth a pound of cure, Russell usefully points out that it is an empirical matter as to whether an ounce of prevention is worth a pound or a gram of cure.*

some sense of the value of what prevention efforts accomplish if resources are to be sensibly allocated. For instance, if prevention efforts result in fewer lives lost in alcohol-related motor vehicle accidents, then something of considerable value is produced despite either the practical or the ethical difficulties involved in actually quantifying such value. Where dollar valuation is not inherently so tenuous (although, as we shall see, still far from simple), the current wisdom (see US DHHS 1990) is that detrimental effects of drinking on performance in the labor market may be the single most important component of alcohol-related "economic costs."

Indeed, considerable casual empiricism about the effects of drinking on work habits has provided researchers in this area with a rich set of hypotheses on the relationships between drinking and productivity, efficiency, absenteeism, earnings, wages, work effort, and other measurable labor market outcomes. The main objectives of this paper are to survey what has been done so far in addressing these questions and to identify directions where future efforts might be most informative.

Shaw's quote in the epigraph captures several key themes of this paper. First, and perhaps most important, is that while the channels through which the use or abuse of alcoholic beverages might affect productivity and work performance are diverse, empirical measurement of their importance is a formidable task. Not only is it the case that most datasets that have been used in such endeavors were not designed for the tasks at hand but, as Shaw notes, many tasks in many occupations may be virtually unaffected by drinking. Since the subtleties of what workers do and what they are expected to do when they are on the job are considerable, quantifying relationships between alcohol use and labor market performance is likely to be quite complex in many situations.

In other settings, however, the effects of drinking on productivity and job performance are far from subtle. The recent Exxon Valdez and New York City subway calamities, for instance, have been attributed to on-the-job inebriation. Accounting quantitatively for the massive property damage and the loss of life and limb involved in these and other perhaps less newsworthy incidents is no easy task. Nonetheless, it remains true that the "product" valued in the marketplace and produced by ship captains and subway motormen is safe transport of cargo and passengers. As such, the consequences of alcohol-related on-the-job accidents are at least in principle relevant to consider in assessing relationships between productivity and drinking even though they may be tough to quantify.[3]

Yet another consideration is that problem drinking can affect labor market success in diverse ways in a given population. One group of individuals may find that problem drinking affects its labor supply or labor market participation;

[3]*This survey does not consider in any detail the issue of drinking and its relationship to job-related accidents. For additional discussion see US DHHS (1987) and Hingson et al. (1985).*

another may have had its occupational options restricted by drinking problems; yet another may be less (or, as some of the results surveyed below suggest, more) productive on the job as a consequence of its drinking behavior; a fourth group may find that its labor market success has been hindered by reduced human capital accumulation. From an empirical perspective, such heterogeneity would imply that there may not simply be a model describing each outcome of interest, but rather that the parameters relating, for example, alcoholism to earnings may themselves vary in the population.

While this point will be addressed in greater detail in a later section, it is worth noting here, if for no other reason than because ostensibly noninformative or insignificant empirical results may in part be due to the failure to recognize *where* the effects of problem drinking are being manifested. For instance, it may be that alcoholic and nonalcoholic shop assistants and billiard markers command the same wage rates and realize the same annual earnings. Yet, if the typical alcoholic shop assistant or billiard marker would have wanted and been able to be a neurosurgeon were he or she not plagued by alcoholism, then it is readily apparent that a comparison of alcoholic and nonalcoholic shop assistants' or billiard markers' earnings does not offer much in the way of guidance as to the costs of alcoholism in such populations.

One central issue to be discussed below is whether alcoholism or alcohol use have causal effects on labor market outcomes. The clinical trials literature on treatment effects provides a standard framework in which to assess such issues.[4] All else equal, does some treatment "cause" a particular outcome to occur relative to what would have happened in the absence of that treatment? It is worth noting at the outset that the inferences made and conclusions reached in most of the literature surveyed here are based by and large on nonexperimental data and are, therefore, most appropriately interpreted as associational rather than causal. Certain econometric identifying assumptions[5] can indeed lend a causal flavor to such results, but the bottom line is that the legitimacy of such identifying restrictions is often tenuous.

This chapter has two main purposes. The first is to provide a survey of the existing empirical literature on the relationships between alcohol use and labor market success. While the preponderance of the empirical evidence to date is on outcomes like income, earnings, and wages, there has of late been increasing attention paid to other labor market phe-

[4] *See Mullahy and Manning (in press) for detailed discussion. Much of modern statistical analysis views the issue of cause and effect in the context of a model developed by Rubin (1974). At the core of Rubin's model is the counterfactual idea that a "causal effect" is the difference between the outcomes (y) if the treatment is administered (T = treated) and the outcome if the treatment is not administered (C = control); i.e., the treatment brings about the "causal effect" $y_T - y_C$. The "fundamental problem of causal inference" is that either y_T or y_C can be observed, but not both.*

[5] *For example, the maintained exogeneity of a measure of lifetime alcohol dependence or alcohol abuse in an individual income equation; see Mullahy and Sindelar (1991b).*

nomena like labor supply, participation and unemployment, absenteeism, occupational choice, and human capital. While the emphasis on outcomes that lend themselves easily to use in dollar valuation efforts is perhaps understandable, the literature survey presented in the second section attempts to provide some order to both the "monetary" and the "nonmonetary" evidence.

The chapter's second purpose is to provide some conceptual tools for better appreciating and, where necessary, reconciling the existing body of empirical evidence on these issues. As the literature review will indicate, the role of alcohol in determining labor market success is anything but well understood, and it is submitted that much of the confusion arises simply because different analysts have been studying different issues using different datasets. As such, the third section identifies a set of fundamental theoretical and econometric issues that should be helpful to bear in mind when trying to make sense of the empirical results.

It should be noted that there are many important aspects of alcohol and the labor market this chapter does not attempt to address. Topics like employee assistance programs, health insurance coverage for alcoholism treatment, and workplace drug and alcohol testing[6] are not addressed here. Readers interested in these and other topics should consult US DHHS (1990) and the recent surveys by Rice et al. (1990) and Cook (1991). In addition, the surveys by Killingsworth (1983), Pencavel (1986), and Willis (1986) are good general treatments of labor market phenomena of the kind considered here.

A SURVEY OF RECENT RESEARCH

The discussion here is divided into three subsections. The first surveys the existing literature on the relationships between drinking and earnings, wages, and incomes, this representing the bulk of the research effort to date. The next subsection focuses on other dimensions of labor market behavior (labor supply, unemployment, absenteeism, occupational choice, human capital) whose relationships to drinking behavior have to date been less thoroughly studied than have income/wages/earnings but that are nonetheless essential to address when attempting to assess in full alcohol's role in the labor market. The last subsection is a summary.

Earnings, Wages, and Income

Due presumably to their theoretical linkages to worker productivity, wages, earnings, and income have proven to be the labor market outcomes most studied by economists interested in alcohol's economic costs. While the next section of this chapter will argue that basic textbook relationships like "wage equals value of marginal product" must be applied with considerable care to real-world markets and that measures like "individual income" may indeed be quite removed from any meaningful notion of productivity, it is nonetheless instructive to obtain a

[6] *See Hoyt (1992) and Glazer and Weiss (1992) for interesting discussions of the workplace-testing issue.*

sense of the relationships between alcohol use and these monetary measures of labor market "success" economists have so far unearthed in their empirical work.[7]

It should be stressed here that while "problem drinking" and "alcohol use" are terms used throughout as general descriptors, their applications in any particular study may be vastly different. One study may focus on all levels of alcohol consumption while another may employ measures of chronic alcohol dependence. Considerable care should be exercised in making inferences in such cases where the underlying measurements may be vastly different (see Sindelar, chapter 8).

In one of the earliest attempts to relate alcohol problems in the household to economic success, Berry and Boland (1977) used data from the 1969 Berkeley Social Research Group Survey to relate household incomes to alcohol problems. For this study, alcohol problems were measured as whether or not an alcohol-abusing male resides in the household, and a household was only included in the sample if there is at least one male age 21–59 in residence. The only control variable at the authors' disposal was the male's age.

Berry and Boland found overall that households in which no alcohol-abusing male resides have 22.5 percent higher mean incomes and 24.3 percent higher median incomes than households where alcohol-abusing males reside. There was, however, a pronounced age pattern to these effects. Table 1 displays the results disaggregated by the male's age. While now somewhat dated both descriptively and methodologically, Berry and Boland's results can probably be said to have provided the launching pad for subsequent inquiry in this area.

TABLE 1

Increase in annual income associated with having no alcohol-abusing male residing in the household

Age	Mean % increase	Median % increase
21–29	23.1	24.4
30–39	21.0	22.0
40–49	23.8	26.4
50–59	11.6	14.5

Results from Berry and Boland (1977) analysis of the Berkeley Social Research Group data.

In another early study whose descriptive usefulness is also probably limited, Benham and Benham (1982) employed data on a sample of males who were children in St. Louis in the 1910's and 1920's and either were referred to a guidance clinic between 1924 and 1929 or were selected as controls. These individuals were followed up 30 years after they were

[7] *The basic idea is that in a competitive marketplace with profit-maximizing firms and utility-maximizing workers, the wage a worker receives will reflect the value that the marketplace places on the additional production of goods and services that the worker brings about for the firm. If workers are less productive because, say, of some health problem, their labor effort will elicit less production and—at least in theory—they should be paid a correspondingly lower wage. Observing the difference in wages between productive (e.g., healthy) and less-productive (e.g., unhealthy) workers permits one in this context to infer something about the productivity foregone as a consequence of the health problem, when all other determinants are appropriately held constant.*

initially surveyed as children. The alcoholism measure used by Benham and Benham was a dummy variable indicating whether after age 18 the individual had a well-established addiction to alcohol without any other major psychiatric disorder (except neuroses). This stringent definition resulted in a 3- to 5-percent lifetime incidence rate for males.

In simple comparisons, alcoholics with positive earnings are found to have mean weekly earnings of $99 while mean weekly earnings for the group of psychiatrically "well" individuals and for nonalcoholics are $119 and $118, respectively;[8] these translate into about a 16- to 17-percent earnings reduction associated with alcoholism in both instances. In a regression setting where other key physical and mental health status variables are controlled, the effects of alcoholism on earnings are negative but statistically insignificant. Depending on what other variables are included in the specifications, the point estimates suggest percent reductions in weekly earnings due to alcoholism ranging from 5 to 11 percent.

In probably the best-known study of the economic costs of alcohol-related problems (commonly known as the RTI study), Harwood et al. (1984) used data from the 1979 National Survey of Attitudes and Interests in Drinking Practices and Problems ("National Alcohol Survey") to assess relationships between household income from all sources and various measures of drinking (quantity/frequency) and problem drinking. An important constraint of this dataset is that only the respondent's drinking status was ascertained in the survey.

The key result from the regression analysis is that problem drinking is associated with a 21-percent reduction in household income from all sources. In addition, the estimated quantity/frequency profile, specified by Harwood et al. as a quadratic response, is concave, suggesting that low-level consumption is beneficial while consumption beyond about 2.0–2.6 ounces per day is harmful.[9] The fact that the problem drinking and quantity/frequency measures are included in the same regressions suggests some concern about the possible effects of multicollinearity on the stability of these results.

For a number of reasons—some conceptual, some econometric—Heien and Pittman (1989) were critical of the Harwood et al. results. Using (as do Harwood et al.) the 1979 National Alcohol Survey data on household incomes and various measures of the quantity and frequency of alcohol consumption and of problem drinking, Heien and Pittman found in a raw comparison that households in which no problem drinker resides have incomes 13.6 percent lower than households that include at least one prob-

[8] *Benham and Benham did not provide the descriptive statistics for nonalcoholics, but the data presented were sufficient to back up these figures.*

[9] *The methods used by Harwood et al. to compute point estimates of the drinking and/or problem drinking effects have been discussed and criticized by several other authors; see Cook (1991), Heien and Pittman (1989), and Rice et al. (1990) for additional discussion.*

TABLE 2

Reduction in annual income due to DSMALC

Age	Males %	Males $	Females %	Females $
18–24	1.4	131	0.8	75
25–34	3.0	807	2.8	484
35–54	5.5	1,958	11.9	2,141
55–64	9.3	2,295	18.7	1,900

Results from Rice et al. (1990) analysis of the ECA data.

lem drinker. In a regression setting, they found—again contrary to Harwood et al.—no statistically significant relationships between household incomes and either the quantity/frequency measures or the problem drinking measures. Since Heien and Pittman were unable to replicate the Harwood et al. results, some caution should probably be exercised in interpreting the latter results.[10]

A recent study by Rice et al. (1990), whose main objective was to update and revise the Harwood/RTI economic cost estimates, contains some results that are useful for purposes at hand. As a component of their economic cost computations, Rice et al. estimated a set of relationships between income and disorders including alcoholism. The authors used multiple site data from the Epidemiologic Catchment Area (ECA) surveys in conjunction with the DSM-III diagnosis of ever having (i.e., lifetime) alcohol abuse or dependence (DSMALC). The income measure used in this study is personal income.

Rice et al.'s key finding was that lifetime alcohol dependence or alcohol abuse is negatively related to personal income for males and females. A summary of the percentage reductions in income associated with DSMALC by age is shown in table 2 (reproduced from Rice et al.).

Several potential concerns about the Rice et al. estimates that cannot be resolved by reference to the text of their published report should be noted at this point. For instance, the authors noted that they utilized the ECA sampling weights in estimation, but it is not clear whether such weights are appropriate for the estimation methods they used. It is also not apparent from the discussion whether the alcoholism-related income losses they reported (i.e., those summarized in table 2) took proper account of the censored nature of their income variable.

Berger and Leigh (1988) used data from the 1972–73 Quality of Employment Survey (QES) to assess the relationships between alcohol consumption and wages. Their empirical approach was to first specify and estimate a selection equation to separate drinkers from nondrinkers and then to estimate selection-corrected wage models for drinkers and nondrinkers. For this analysis drinkers were

[10]*It might be noted that Rice et al. (1990) share the same concern and also point out some potential problems with the Heien-Pittman results. In private communication, Dale Heien has informed me that some of the Heien-Pittman methodology was not appropriately reported in the Rice study. Given the considerable lapse of time that has occurred, a full reconciliation of these apparently disparate results is unlikely to occur.*

TABLE 3

Percent that nondrinkers' wages are lower than drinkers' wages

Calculation method	Males	Females
I	12.8	25.5
II	7.8	20.9
III	36.2	28.7

Results from Berger and Leigh (1988) analysis of the QES data.

defined as individuals reporting drinking alcoholic beverages at least one or two times per week. The key sample selection criteria were that the individual had to be at least 18 years old and working for pay at least 20 hours per week; separate estimates were obtained for males and females. Given the limitations of the dataset, self-selection into the sample of workers was not controlled for; the implications of this limitation are discussed below.

Table 3 summarizes the key findings of the Berger-Leigh study. Since the authors suggested at least three different methods that might be used to compute a "wage differential," results are shown for each method. As evident from the table, even the most conservative approach to estimating wage differentials (method II) suggests sizable wage differences between drinkers and nondrinkers. It is important to stress, however, that the definition of "drinkers" employed by Berger and Leigh is quite different than any reasonable definition of "problem drinkers." Since their "drinkers" covers the spectrum from serious problem drinkers to very light drinkers, measurement error or functional form misspecification effects on the parameter estimates may be important to consider.

In his comprehensive survey of the social costs of drinking, Cook (1991) attempted to replicate the Berger-Leigh results using the QES data set described above. Table 4 reproduces the results reported in Cook's figure 2.1. The most striking aspect of these results is that they basically confirm the Berger-Leigh findings. While Cook's statistical analysis is more direct than the Berger-Leigh selection-correction methodology, essentially the same message emerges: At least over some range of moderate drinking, workers' earnings increase with alcohol consumption.

In light of these results, it should be emphasized that categories 1 through 5 represent light and moderate drinkers (averaging less than one drink per day), categories 6 and 7 represent a somewhat higher level of consumption (averaging two to four drinks per day), and category 8 represents the heaviest drinkers (averaging four or more drinks per day). The definition of category 8 is such that it includes approximately the top decile of the distribution that is often the focus of policy discussions. If the individuals in this category are the true "problem drinkers," then more attention ought to be paid to the structure of the alcohol/earnings relationship in this tail of the distribution. For instance, it would be useful to know the within-category earnings variance and how this changes over income levels; if the relatively large median in cat-

TABLE 4

Median annual earnings by drinking status

Alcohol consumption	Median annual earnings
1. Never drink (n = 98, c.f. = 0.12)	$9,000
2. Less than one drink per year (n = 52, c.f. = 0.19)	$8,500
3. Less than one drink per month (n = 78, c.f. = 0.29)	$9,700
4. 1–5 Drinks per month (n = 93, c.f. = 0.40)	$10,900
5. 6–29 Drinks per month (n = 203, c.f. = 0.66)	$11,000
6. 30–59 Drinks per month (n = 116, c.f. = 0.80)	$10,000
7. 60–119 Drinks per month (n = 87, c.f. = 0.91)	$10,000
8. 120 Drinks per month or more (n = 71, c.f. = 1.00)	$10,400

Results from Cook (1991) analysis of the QES data.
n, number of observations in each category; c.f., sample cumulative frequency.

egory 8 masked a relatively large variance in earnings, this would itself be an interesting result.

Indeed, the key relationship Cook estimated is a positive relationship between alcohol consumption and earnings up to the level of one drink per day (category 4) that, unless concentrated on few drinking occasions, would typically not constitute problem drinking behavior. No statistically significant trend was found beyond category 4. It should be stressed, of course, that not all levels of positive consumption in Cook's data can be considered as heavy drinking; that is, the distinction between "drinkers" and "nondrinkers" in this data should in no way be meaningfully viewed in the same way as a distinction between "nonproblem drinkers" and "problem drinkers" or between "light drinkers" and "heavy drinkers."[11]

In a series of recent papers, Mullahy and Sindelar utilized data on males from the ECA survey's New Haven site to analyze a variety of relationships between alcoholism and labor market success; Mullahy and Sindelar (1991*b*) summarized the key results from this work insofar as income is concerned. The income measure used by Mullahy and Sindelar is the individual's income "brought in" from all sources, which includes (presumably) labor market earnings, interest and dividends, and transfer payments. Mullahy and Sindelar considered several measures of alcoholism problems, but their key results are based on the DSM-III measure of ever having met criteria for alcohol dependence and/or alcohol abuse.

[11] *Using data from the 1984 National Alcohol Survey, Model (1992) also found a positive relationship between the frequency of alcohol use and wages.*

The results based on simple comparisons of mean income differences between nonalcoholics and alcoholics are summarized in table 5.

Several findings are striking. First is that there is a pronounced age pattern to the differences: The sizable negative effects of alcoholism apparent in the pooled data owe largely to the "prime working age" subsample (ages 30–59) which are partly offset by the positive effects seen in the youngest and oldest subgroups (compare with table 1). Second is that the inferences one might draw about the effects of alcoholism depend importantly on whether all individuals or just full-time workers are being studied. The fact that alcoholism's negative effects are smaller and less significant for workers than for all individuals (at least in the prime working ages) suggests that alcoholism may have nontrivial

TABLE 5

Individual income (in thousands of dollars) by lifetime alcoholism status

Ages	All observations		Full-time workers only	
	Nonalcoholic	Alcoholic	Nonalcoholic	Alcoholic
All ages	20.8	17.4	23.2	20.7
	(712)	(185)	(552)	(133)
	{-16.3%}***		{–10.8%}	
Age subgroups				
22–29	12.8	13.2	15.7	15.4
	(178)	(65)	(116)	(48)
	{+3.1%}		{–1.9%}	
30–44	24.2	21.1	25.2	25.3
	(273)	(75)	(239)	(55)
	{–12.8%}		{+0.4%}	
45–59	24.7	16.8	25.7	19.6
	(169)	(38)	(145)	(26)
	{–32.0%}***		{–23.7%}***	
60–64	18.8	20.6	23.4	29.4
	(92)	(7)	(52)	(4)
	{+9.6%}		{+25.6%}	

Results from Mullahy and Sindelar (1991b) analysis of New Haven ECA males sample.
*For two-tailed tests of differences in means, allowing for different variances, * = $p<0.10$, ** = $p<0.05$, and *** = $p<0.01$.*
Number of observations in parentheses; decrease or increase for alcoholic relative to nonalcoholic in braces.

impacts on labor supply behavior; this issue is taken up in the next subsection.

Essentially the same results carry over to a regression setting where Mullahy and Sindelar focused on the 30–59 age group. The key findings from that analysis are that alcoholism has a negative and significant relationship with individual income controlling for other covariates, but that the size and the significance of the relationship depends crucially on what other covariates are controlled. When only a sparse set of controls (age and race) is included, alcoholism is estimated to result in a 31-percent reduction in mean income; when the full set of controls (including schooling, marital status, physical health status, and other household income) is used, the estimated effect of alcoholism shrinks to 17 percent.

This sequence of results led Mullahy and Sindelar to conclude that there may be both direct and indirect effects of alcoholism on income and earnings. Put simply, the direct effects are those captured by regression coefficients, while the indirect effects are those manifested by, for example, youth problem drinking retarding schooling attainment, reduced schooling attainment affecting earnings, and drinking problems in youth carrying over to adult alcoholism. This issue is considered in Mullahy and Sindelar (1990*a*,*b*), and is addressed below in the section entitled "Some Basic Issues."

Finally, it is interesting to note that while Bartel and Taubman (1979)—one of the seminal contributions to the health-earnings literature—did not specifically consider alcoholism as a component of health capital, they did include a measure of liver, gall bladder, and pancreas disease. In one key specification, this indicator was found to have a significant negative effect (33-percent reduction) on the earnings of males approximately 50 years old. To the extent that this disease measure is correlated with prior chronic problem drinking (see US DHHS 1990, chapter V), then it is reasonable to suggest that some share of these earnings reductions would be due to the alcohol problems themselves.

In closing this subsection, one potentially important complication involved in analyzing income data should be noted, that is, the implications for individual income measurement of transfer payments and income earned or generated by other household members. It would be ludicrous to argue, for instance, that higher levels of transfer payments represent, all else equal, greater productivity, since eligibility for and receipt of public transfer payments often coincide with subpar productivity. However, to the extent that measures of individual or household income include transfers received by the individual or household, then precisely the "wrong" signal is being sent by reference to such aggregated income data. Analogous, albeit somewhat more complex, arguments apply when analyzing household income measures and recognizing that household incomes may, all else equal, increase due to the alcohol problems of spouse one when spouse two increases work effort to compensate for the income effect associated with spouse

one's alcohol-related reduced earnings. Empirical analysis of such labor supply responses is both conceptually challenging and demanding of data, although the work by Berger and Fleisher (1984) provides one conceptual template for studying such responses.[12,13]

Other Facets of the Labor Market

The studies surveyed in the previous subsection are anything but definitive in establishing whether drinking or problem drinking results in labor market problems when these are measured by income, wages, or earnings. The conceptual analysis presented below in the next section suggests, however, that not only are distinctions between income, earnings, and wages important in addressing such questions, but that focusing on these measures alone may divert attention from other important dimensions of labor market behavior that must also be understood if some order is to be made of the seemingly disparate results appearing in the literature. Unfortunately, compared with what is admittedly a relatively small body of evidence on the links between problem drinking and income-earnings-wages, the evidence on other dimensions of the alcohol-labor phenomenon is even more limited.

Labor supply, labor market participation, and employment

Despite the obvious importance of labor supply factors in the determination of wages, earnings, and income, and the reasonable hypothesis that problem drinking might be an important influence in individual and household labor supply decisions, to date disappointingly little attention has been paid to such matters in empirical work. While no evidence could be turned up on how drinking relates to hours and weeks of work, there is some

---

[12]Berger and Fleisher found the relationship between a wife's labor supply and a husband's poor health to depend on the generosity of transfers that would be available as a consequence of the husband's ill health. In effect, the wife's market earnings and transfers function like substitute inputs in the production of household incomes. While the Berger-Fleisher results are intuitively plausible, one potential reservation about their applicability to alcohol-related health problems owes to the possibility of alcohol-related assortative mating, e.g., the possibility that husbands with alcohol problems are relatively more likely to have wives with alcohol problems. Were this true, then determining spouse one's labor supply response to spouse two's health problem obviously must take into account the effects of spouse one's alcohol problem on his/her own labor supply behavior. This would be an extremely interesting avenue for future research to consider.

[13]Using the subsample of males aged 30–59 from the New Haven ECA data described above, I estimated a set of probit models for the binary outcome of receipt of four categories of public transfers: disability payments or benefits; unemployment compensation; Social Security benefits other than disability (probably supplemental security income, given the age composition of the sample); and welfare payments (probably general assistance, as the sample is males only). For the models specified as a function of lifetime alcoholism (dependence and/or abuse), the alcoholism coefficient estimate is positive and significant for welfare receipt and for Social Security benefits other than disability. None of the models specified as a function of 1-year alcoholism (again, dependence and/or abuse) resulted in a statistically significant coefficient estimate for that variable. These results are admittedly preliminary and are intended only to be suggestive; further research on this issue would be useful.

emerging support for various hypotheses about how labor market participation and (un)employment are affected by alcohol-related problems.[14]

Labor market analysts typically find it useful to distinguish three states of labor market participation: currently employed, currently unemployed, and out of the labor market. Individuals in the first two categories are considered labor market participants while those in the third category are nonparticipants. Not all surveys offer information based on these standardized categories, however, and other definitions of labor market participation or attachment also are in use. Some care should thus be exercised when comparing results across different datasets.

Using the St. Louis males data described above, Benham and Benham (1982) also studied the relationship between alcoholism and employment. By their standards, an individual was considered employed if he was a full-time employee, a part-time employee, or in the military at the time of the followup survey. The simple comparisons indicate that alcoholics have an overall employment rate of 86 percent while the rate for the sample of psychiatrically "well" individuals and for the sample of nonalcoholics are 95 percent and 84 percent, respectively. For full-time employment, the respective rates are 79 percent for alcoholics versus 90 percent for the well group and 78 percent for nonalcoholics. The fact that alcoholics have slightly (although not significantly) higher rates of employment than nonalcoholics is somewhat perplexing, but the considerable differences between the composition of the "well" group and the nonalcoholic group suggest that many nonalcoholics otherwise are not well. In regression models that control for other components of human capital, inferences about the relationships between alcoholism and employment (full- or part-time or military) were mixed: sometimes positive, sometimes negative, but in no instance even close to statistically significant.

Mullahy and Sindelar (1991*b*) used a criterion for employment that was far more stringent than that used by Benham and Benham: With the criterion for full-time employment being that the individual worked in all 12 months of the preceding year with the possible exception of paid vacations, Mullahy and Sindelar found for the New Haven ECA sample of males ages 30–59 (described above) that ever having met the DSM criteria for alcohol dependence or abuse is negatively and significantly related to full-time work propensity. The Mullahy-Sindelar results are summarized in table 6.

The relatively small differences between alcoholics and nonalcoholics when pooled over all age groups mask striking differences seen for the specific age groups. As seen in table 6, for males ages 30–44 and 45–59, the differences between the full-time work propensities of

[14]*Two key early contributions to the health and labor supply literature that are worth consulting are Berkowitz and Johnson (1974) and Parsons (1977).*

TABLE 6

Full-time employment status by lifetime alcoholism status

Ages	Alcoholism status: Nonalcoholic		Alcoholic
All ages	0.775		0.719
	(712)		(185)
		{–0.056}	
Age subgroups			
22–29	0.652		0.738
	(178)		(65)
		{+0.086}	
30–44	0.875		0.733
	(273)		(75)
		{–0.142}***	
45–59	0.858		0.684
	(169)		(38)
		{–0.174}**	
60–64	0.565		0.571
	(92)		(7)
		{+0.006}	

Results from Mullahy and Sindelar (1991b) analysis of New Haven ECA males sample.
*For two-tailed tests of differences in proportions between ALC=0 and ALC=1, * = p<0.10, ** = p<0.05, and *** = p<0.01.*
Number of observations in parentheses; decrease or increase due to alcoholic relative to nonalcoholic in braces.

alcoholics and nonalcoholics are significant. These significant negative effects of alcoholism carry over to a regression setting where age, race, and other covariates are controlled. Mullahy and Sindelar ascribe at least part of the positive (albeit insignificant) relationship between alcoholism and employment in the youngest age group (22–29 years) to the possibility that many individuals of this age are still in school and that individuals may be more likely to still be in school if they are not affected by problem drinking.[15]

Despite the markedly different definitions of alcoholism and employment used by Benham and Benham and by Mullahy and Sindelar, the results in both cases suggest major employment problems associated with alcoholism for males in the prime working years—at least to the extent that they are not otherwise unhealthy.[16]

Absenteeism

Another aspect of the labor supply decision that has attracted attention is work loss. To date, perhaps the most useful national estimates of the relationships between drinking and work loss are those obtained by Manning et al. (1991). These

[15]*This could obviously cut both ways. While problem drinking may impede someone from, for example, pursuing graduate studies, it could also prolong academic careers by hindering progress toward a B.A. degree. Additional discussion of the relationships between problem drinking and schooling is provided in the third subsection.*

[16]*Mullahy and Sindelar found that good physical health is a positive and significant contributor to full-time employment propensity, thus reconciling some of the apparent differences between Benham and Benham's raw differences and those obtained by Mullahy and Sindelar. Also, while the Benham and Benham study does not provide age data, the sample mean age was 43 years, with a presumably fairly small variance due to the structure of the initial survey's design.*

researchers used data from the Rand Health Insurance Experiment (HIE) and from the 1983 National Health Interview Survey (NHIS) to address how work loss (as a dependent variable) is related to various categorizations of alcohol use.

The Manning et al. results from their analysis of the HIE data indicated that former drinkers have 38 percent more work loss compared to abstainers and infrequent current drinkers. However, Manning et al. uncovered no statistically significant relationships for current drinkers between the monthly volume of reported alcohol consumption and the amount of work loss. The NHIS data yielded somewhat different conclusions: Like the HIE results, the NHIS analysis indicated no significant relationships between quantity consumed by drinkers and work loss; however, the significant former drinker results found in the HIE analysis were not replicated in the NHIS analysis.

It should be stressed that the sample composition caveats raised above about the Berger-Leigh and Cook studies apply with equal force in the analysis of absenteeism or work loss. Simply put, individuals don't lose time from work unless they already have been working. For instance, the skip logic of the Health Interview Survey is such that the work loss question is asked only of individuals who satisfy the NHIS criterion for "currently working."[17] Accordingly, should problem drinking's most consequential effect be on whether one works or not, then it would not be surprising to see little or no relationship in the data between work loss and alcohol-related problems.

Human capital

As will be discussed in greater detail in the third section of this chapter, the picture of how drinking relates to labor market success is necessarily incomplete unless the fundamental role of human capital is included. Economists have long held that productivity and success in the labor market are largely determined by the level of productive human capital an individual brings to the workplace.

The most obvious measurable components of human capital are the individual's schooling attainment, work experience (both general and job-specific), and health status. However, other factors—measurable or not— like the quality of one's schooling, one's upbringing, and one's family structure and marital status are also recognized as potentially productive elements of one's productive human capital stock.

For present purposes perhaps the most important consideration is that if, as suggested above, problem drinking as a youth or young adult retards the accumulation of human capital, then when examining relationships between alcohol problems and labor market success several potentially important pathways must be considered. If the hypothesis is correct, then while adult problem drinking may

[17] *By NHIS standards, "currently employed" means (a) worked in past 2 weeks; or (b) did not work in past 2 weeks, has job, not on layoff, not looking for work; or (c) did not work in past 2 weeks, has job, looking for work. It should also be pointed out that the "work loss" is work loss due to some condition.*

have direct negative consequences for labor market success, the fact that adult problem drinking is surely positively correlated with youth problem drinking suggests that these adult problem drinkers will probably have relatively low levels of productive human capital as well.

At this juncture, possibly the best that can be said of what is known about the relationships between drinking and human capital is that there is now at least the recognition that this is an extremely important area to assess empirically. The literature to date contains only a handful of suggestive results that will be discussed here to provide readers with a flavor of the direction in which this research is proceeding.

The recent well-publicized emphasis by Surgeon General Antonia Novello on youth problem drinking underscores the concern about this issue in the public policy sphere. To date, unfortunately, still very little is known about how youth drinking affects the level of schooling attainment and its quality. Mullahy and Sindelar's (1989, 1990*a*) results suggest that onset of alcoholism symptoms during youth is associated with reduced schooling attainment; specifically, first onset of alcoholism symptoms before age 19 is found to be related to an 11-percent reduction in schooling attainment controlling for other covariates but only an 8-percent raw difference. It should be stressed, however, that the direction of causality between alcoholism symptoms and schooling cannot be ascertained with confidence in this study.

The Benham and Benham (1982) study described earlier also provides some raw comparisons of schooling attainment between alcoholics and "well" individuals. Recalling from above that the Benham-Benham definition of alcoholism is quite strict, alcoholics were found to have mean schooling attainment 24 percent and 15 percent below that of the "well" subsample and the nonalcoholic subsample, respectively. Given the considerable difference in the definition of alcoholism used in their studies (recall that Mullahy and Sindelar's measure was onset prior to age 19), the 11-percent and 15-percent differences found in the two studies are surprisingly similar.

A recent study by Cook and Moore (1992) attempted to determine whether and, if so, how youth drinking is related to educational attainment. Cook and Moore used data from the National Longitudinal Survey of Youth (NLSY) to test various hypotheses about how youth drinking might influence educational attainment.

The fundamental message of the Cook and Moore results was that drinking impedes educational attainment. Students who were frequently drunk (at least four occasions of consuming six or more drinks in the past month) during their senior year of high school had significantly lower propensities to matriculate to college and graduate with 4-year degrees, and had lower overall years of completed schooling than nondrinkers or more moderate drinkers. For those in the "frequently drunk" category, the proportions matriculating and graduating were 35 percent and 14 percent, as opposed to 42 percent and 23 percent among those not frequently drunk. In addition, the

level of alcohol consumption is itself a negative predictor of matriculation, graduation, and years of completed schooling.

Occupational choice

To date there has been little empirical analysis of the relationships between problem drinking and occupational choices.[18] Conceptually, many of the issues germane to understanding how human capital accumulation relates to alcohol use are relevant in this context as well, since occupational choices or options are in large measure determined by the quantity and quality of human capital that an individual brings to the labor market. Mullahy and Sindelar (1989) also suggested that rational job search behavior would require job seekers to weigh all components of their compensation package, not simply wages and salaries. It may be the case, for example, that jobs offering relatively generous sick leave, flexible hours, and health insurance but relatively low wages are more attractive to individuals with health problems (e.g., alcoholism) than jobs offering less attractive fringes but higher wages. If this is true, then measuring productivity effects of alcoholism by reference to wage differentials will tend to overstate the magnitudes of such effects, since the entire value of the compensation package is not accounted for. This issue is considered in greater detail below.

Summing Up

It would be gratifying to report that the existing empirical literature on the relationships between alcohol and labor market outcomes has attained solid consensus on the magnitudes and signs of such relationships. As this survey demonstrates, however, the literature has not attained consensus. Yet this conclusion should be interpreted in light of two fundamental realities.

The first is simply that the body of existing literature is by any reasonable standard still quite small. The second is lack of consensus in measures used in empirical work. This survey has turned up no instance where a study has attempted to replicate the results of a previous study using the same outcome measure (income, wage, participation, etc.) and the same alcohol measure (2-week ethanol consumption, lifetime alcohol dependence, etc.) but different data. That is, the lack of consensus is due more to the fact that researchers have actually been exploring somewhat different hypotheses using different sets of data than to the failure to replicate earlier findings.

In this light, some key distinctions the reader might bear in mind when assessing this literature include (but are certainly not limited to) the following: (1) whether workers or all individuals were sampled;[19] (2) whether males or females were sampled; (3) whether income, earnings,

[18] *Plant (1977) provides a survey of alcoholism rates in various occupations. It might be noted at this juncture that even the general labor economics literature on occupational choice is not large; see Mullahy and Sindelar (1989) for a survey and for some estimates of the relationships between alcoholism and occupational status.*

wages, or some other measure was the focus; (4) whether alcohol consumption or alcoholism was the explanatory variable; (5) whether the age composition of the sample was homogenous or heterogeneous; and (6) whether other covariates that might be correlated with alcohol consumption or alcoholism were included as controls.[20]

On a more positive note, what the existing research has fruitfully accomplished is essentially a delineation of fundamental issues that ought to be considered in future research efforts. Indeed, this is probably all that can reasonably be asked of a relatively young (with a couple of exceptions, less than a decade old) and relatively small field of inquiry.

SOME BASIC ISSUES

Why should we be studying relationships between drinking and labor market success? What can we hope to learn from econometric analyses of such relationships? How can the statistical estimates obtained in such research be used fruitfully in policy debates? While perhaps more suitably posed at the outset, these and other fundamental questions can no longer be neglected. The survey offered above indicated, if nothing else, that there is a growing literature on the relationships between alcohol and labor market outcomes. The purpose of this section is to explore some issues that are crucial in interpreting such empirical research (extant and future) and hopefully, as a byproduct, to suggest some avenues for future empirical investigations.

Implicit thus far in much of the discussion have been microeconomic models of human capital accumulation and labor supply. Space does not permit an exposition of such models here, but as a point of reference the basic themes in human capital are those developed by Mincer (1974), Becker (1975), Grossman (1972), and Mushkin (1962), with important early applications to the earnings-health capital issue by Bartel and Taubman (1979, 1986) and Luft (1975) (see Willis 1986 for a good survey of this literature). Fundamental work in labor supply and wages is summarized nicely by Deaton and Muellbauer (1980), Killingsworth (1983), Pencavel (1986), and Smith (1980).

This section is divided into two subsections. The first subsection assesses some more-or-less theoretical or conceptual issues, while the second subsection

[19] *At this juncture the reader should be reminded that a key consideration in drawing inferences from the Berger-Leigh and Cook results is that their samples consisted only of workers. Interestingly, Kaestner (1991) and Gill and Michaels (1992) found use of illicit drugs to be positively related to wages. Like Berger-Leigh and Cook, Kaestner's analysis is effectively confined to workers. Gill and Michaels (1992) found that users of any illicit drug(s) had lower employment rates than nonusers, but that hard drug users did not show differing rates. Register and Williams (1992) found negative employment effects of marijuana use.*

[20] *It also never hurts to recall that there are two kinds of error possible in statistical inference, and the lack of attention to test power can be particularly misleading if sample sizes are small or if samples are not particularly well-designed for the purposes at hand. See Andrews (1989) for an exceptionally lucid discussion of power in econometric work.*

considers some econometric and statistical topics. The discussion here does not attempt to be at all exhaustive, and there are surely other topics that are at least as important to address, but those issues surveyed here should provide a baseline for future explorations.

Theoretical and Conceptual Issues

This subsection has three components. The first is a description of a simple accounting framework that might be useful in thinking about many of the data issues encountered above. The second is a discussion of the relationships between wages, productivity, and welfare. The third is an assessment of the possible role of indirect effects of problem drinking on labor market success.

Problem drinking and the labor market: A taxonomy

It was suggested earlier that, due to individual or market idiosyncracies, alcohol-related problems may be manifested in the labor market in quite different ways. A useful interpretation of the literature surveyed above is that each study has considered some element or elements of the big picture, but that there has been no real attempt to assemble in any coherent way all the pieces. Accordingly, it seems useful to construct a simple model of labor market outcomes for assessing and, to some degree, reconciling some of the empirical results surveyed above.

As a starting point, note that a household's pretax income at a given point in time (y) can be described approximately by the primary breadwinner's labor earnings plus income earned by other household members plus nonlabor income (dividends, interest, etc.) plus transfer payments, that is,

$$y = w_j L + H + N + T$$

where w_j is the wage or salary rate of the primary breadwinner in job j, L is the amount of labor supplied by this individual, H is other household members' labor income, N is nonlabor income, and T is transfer payments.[21]

This simple accounting relationship is useful for purposes at hand once it is recognized that all determinants of household income may be affected by a primary breadwinner's problem drinking. That is,

$$y = y(A,.) = w_{j(A,.)}(A,.)L(A,.) + H(A,.) + N(A,.) + T(A,.)$$

where A is an indicator of problem drinking. This framework makes explicit the fact that predictions about how drinking affects income or any component(s) of income will in general depend on several responses. It is, for instance, certainly plausible that H and T could be positively related to A even if the primary breadwin-

[21] *A more thorough description of the implications of alcohol-related problems for the household's budget set would be obtained by factoring in in-kind transfers received by the household (e.g., food stamps, housing allowances, Medicaid, etc.), whose receipt and/or level would also presumably be influenced by alcohol-related problems. The simpler framework described here will suffice for the present discussion, however.*

ner's labor earnings are negatively related to A.

However, even focusing specifically on individual labor earnings, $w_{j(A)}(A,.)L(A,.)$, one sees that there are at least three key dimensions to consider in predicting the effects of drinking on earnings. The occupation index, j, is specified to depend on A since, as argued earlier, the choice of occupation itself plausibly depends on one's drinking status.[22] Conditional on working in job j, the wage rate earned may itself depend on drinking status. Finally, the amount of labor actually supplied to job j is likely to depend on, or at least be correlated with, drinking behavior. Predicting the direction of the effect of drinking behavior on labor market earnings thus requires consideration of three partial derivatives, and a priori it is not at all evident that all three are negative.

Among other things, this framework suggests that observed wages may depend on drinking status. A very simple model of wage determination and labor market participation is one where the potential participant compares his or her reservation wage with the wage offered by a potential employer. If the reservation wage exceeds the wage offer, then the individual's optimal labor supply is zero; if the wage offer exceeds the reservation wage, then the individual accepts the job and works some positive number of hours.

If information is perfect and employers anticipate that workers with drinking problems will have lower productivity than other workers, then, ceteris paribus, the wage offers extended to problem drinkers will in theory[23] be lower than the wage offers extended to the rest of the population. If the individual's reservation wage is not affected by alcohol use, then this suggests a lower probability of an acceptable wage offer for problem drinkers and, therefore, a higher rate of nonparticipation.

It is equally plausible that an individual's reservation wage will depend on drinking status. Without a more structured model, it is not a priori obvious whether, with all else held constant, problem drinking would increase or decrease reservation wages. Even more formalized models might have a tough time making clear predictions about this sign, since income effects may be important.[24] With imperfect information (the more likely situation), predictions are all the more difficult. The bottom line is that it will generally be difficult to predict whether alcohol-related problems will increase or reduce the propensity to work unless considerable structure is built into such a

[22] *Several of the studies surveyed in the second section included occupation variables as regressors (Berger and Leigh 1988; Mullahy and Sindelar 1989). If it is true that $j = j(A)$, then including occupational measures as rhs variables is perhaps defensible on some grounds, but it also adds another channel through which indirect effects might operate and, more importantly, it also results in a set of potentially endogenous regressors being added to the model.*

[23] *In fact, there may be legal constraints on how much "discrimination" employers can undertake with respect to various components of the employment package (wages, hiring/firing, health insurance provision, etc.).*

[24] *In a similar vein, Phil Cook has suggested to me the possibility that alcoholics, strongly motivated to obtain the means to obtain alcoholic beverages, may work more or work harder than nonalcoholics.*

model. Implications for the realized wages of workers are also fuzzy, although at a minimum one should expect observed wages of workers to demonstrate some variation with drinking behavior when reservation wages, offered wages, or both depend on drinking status.[25]

Understanding the role played by alcohol in the decision to participate in the labor market should be helpful in using the results reported by Berger and Leigh and by Cook on wage effects for workers to draw inferences that might be more broadly applicable. The Mullahy-Sindelar results on the importance of problem drinking as a retardant of full-time work propensity would perhaps be a useful benchmark in such explorations.

Wages, productivity, and welfare

What economically important and interesting information is contained in observed wage and/or productivity differentials? Prominent in policy discussions about the "economic costs" of problem drinking is that (the dollar values of) productivity losses due to alcohol-related problems are useful and interesting to measure. Understanding how drinking affects productivity would seem to be an interesting objective fully apart from the question of whether the dollar value of such productivity losses measures anything interesting from a theoretical perspective. This section, however, focuses on the latter question, that is, what economically useful information is conveyed by understanding the structure of alcohol-related wage differences.

There are two fundamental concerns at issue: productivity and welfare. In a purely technological sense, understanding how productivity (or marginal productivity) is affected by drinking is certainly an interesting question. The beauty (or, perhaps, the curse) of competitive market theory is that it provides the link from such purely technological considerations to welfare considerations by reference to the theoretical result that profit-maximizing firms will pay their workers, at the margin, exactly the amount that the marketplace values their marginal product. The theoretical equality of wage and value of marginal product thus provides a starting point for more elaborate welfare analyses that take into account the value of consumer and supplier surplus.

But how helpful a starting point is this theoretical result? The purpose of this discussion is to suggest several reasons why observed wages may not offer reliable guidance as to the value the marketplace attaches to workers' output.

The first point is obvious. Wages are but one component of an employee's compensation package and, for reasons discussed above, the value of the entire compensation package is the fundamental consideration for both employers and employees. In private industry, wages and salaries accounted for only 73 percent of total compensation in 1988 (Piacentini and Cerino 1990). Thus, even if the theoretical

[25] *The paper by Strauss (1986) on health-related efficiency wages may provide a useful framework in which such issues can be empirically untangled. Consideration of this approach should be on the future research agenda.*

value of marginal product (*vmp*) equality holds, the correct statement of the result is that compensation, not wage, should be equal to *vmp*. Other considerations aside for the moment, this suggests that absolute wage differentials will understate differences in *vmp*, although the percentage differences may be the same.

However, other considerations cannot be cast aside so easily. As suggested earlier, rational job search requires workers to compare the different compensation packages they are offered and choose the job whose compensation package yields the greatest level of utility. If (as hypothesized earlier) problem drinkers differ systematically from those without drinking problems in their preferences about the composition of compensation packages and, on average, prefer compensation packages relatively generous in fringe benefits and relatively less generous in wages,[26] then reference to wage differentials alone will result in an overestimate of the percentage differences in *vmp* and—for reasons suggested in the last paragraph—an ambiguous (though biased) estimate of absolute *vmp* differentials.

A second consideration is that when markets are not competitive, even the theoretical result about compensation being equal to *vmp* is vitiated. The standard result from noncompetitive market theory is that workers will be paid by profit-maximizing firms a wage rate equal to their marginal revenue product (*mrp*), which is generally less than *vmp*. That is, the marketplace values the worker's marginal product more than what the worker is being compensated for producing it, so the previous result on compensation measuring welfare at the margin no longer holds. In addition, the second-order issue of measuring consumer and supplier surplus becomes complicated by the need to account for deadweight losses due to underproduction.

Finally, in addition to the two previous reservations, the theory of efficiency wages provides a final reason why reference to wage (or, even more generally, total compensation) as a standard for measuring the value the marketplace attaches to worker productivity is problematic (see Akerlof and Yellin 1986 for a survey). The theory of efficiency wages suggests, among other things, why wages may not adjust smoothly with changes in marginal product and why it is reasonable to expect that at any point in time the firm's optimal compensation package may not perfectly reflect workers' productivity. Although one might expect some general positive relationship between a worker's productivity and the compensation he or she receives, that is, $c = g(mp)$, efficiency wage theory suggests any number of plausible reasons why the textbook relationship, $c = vmp$, might not be expected to hold for any particular individual at any given point in time.

An altogether distinct issue concerning wages, income, and welfare is considered by Mullahy and Sindelar (1992).

[26] *For example, as discussed earlier it may be the case that jobs offering generous sick leave, hours flexibility, and health insurance for substance abuse treatment but relatively low wages are relatively more attractive to alcoholics than jobs offering less attractive fringes but higher wages.*

Apart from the issues considered above, Mullahy and Sindelar argue that an important shortcoming of standard productivity-based welfare arguments is that they fail to account for real-world ex ante uncertainty in wage and income outcomes. To the extent that individuals are risk averse and to the extent that drinking problems result in a "riskier" (i.e., higher ex ante variance) economic environment, then simply valuing ex post productivity differences will understate from the individual's perspective the welfare losses associated with problem drinking.[27]

Finally, it is interesting to consider the important question of which costs of drinking are internal and which are external. Cook (1991) and Manning et al. (1991) offer excellent discussions of this issue (see also Cook and Moore, this volume). As far as labor market outcomes are concerned, this distinction is clearly of some relevance, particularly when the focus is on whether drinking-related welfare losses realized by other members of the household are internal or external.

For example, if a primary breadwinner's problem drinking reduces his or her earnings, is the fact that this individual's children do not visit museums or attend college for financial reasons of any concern from "society's" perspective? Or is it simply a matter of concern for the family? The answer is not immediately apparent. Yet, despite the fact that such welfare losses are far more subtle than second-party alcohol-related highway fatalities, they may still be of enormous magnitude when extrapolated over the entire population. Whether this is any of "society's" business is an interesting and important philosophical question to ponder.

Direct and indirect effects of problem drinking

Some discussion in the second section hinted that the effects on labor market success of alcohol-related problems may not be fully captured in the estimated regression coefficients for the alcohol variables. To the extent that other determinants of labor market outcomes are themselves determined by problem drinking, then some of the effects of problem drinking on labor market outcomes are indirectly channelled through these other covariates; that is, there may be direct as well as indirect effects of problem drinking on labor market success.

To understand this process, consider the following simple structural model of human capital accumulation and labor market success (the indicated signs are suggestive of plausible partial derivatives):

$$K_2 = f(A_1, K_1, Z)$$
$$\quad - \quad +$$

(human capital accumulation model)

$$A_2 = g(A_1, K_2, Z)$$
$$\quad + \quad -$$

(persistence of problem drinking model)

and

[27] *Indeed, Mullahy and Sindelar (1992) provide some empirical support for the hypothesis that problem drinking does increase the variance of income.*

$$y_2 = h(A_2, K_2, Z)$$
$$\quad - \quad +$$

(adult earnings model)

The first equation indicates that the level of productive human capital available as an adult depends on problem drinking behavior as a youth (A_1), initial endowments of human capital such as family background (K_1), and other exogenous factors (Z). The second equation describes the persistence of problem drinking over time by specifying that adult problem drinking is related to youth problem drinking and other factors. The third equation is the standard human capital earnings model, in which labor market success depends on the level of accumulated productive human capital.

These equations describe in simple terms the notion that there might be important indirect effects of alcoholism on adult labor market success. To see this, view A_1 as a variable that can be affected by some policy instrument and trace through the ultimate effects on adult labor market success of policies that succeed in reducing A_1. Such policies would enhance adult labor market success for two reasons: adult problem drinking would be mitigated (the direct effect) and productive human capital would be enhanced (the indirect effect). The reasonableness of such a model is ultimately an empirical matter, but to the extent that such an account is empirically descriptive, then policies that thwart problem drinking in youth and early adulthood are probably sensible to consider (see Mullahy and Sindelar 1990*a* for further discussion).

Econometric Issues

This section describes a handful of fundamental econometric results that should be useful to bear in mind when interpreting the empirical results surveyed above.

Complications owing to endogeneity and simultaneity

Most of the research surveyed above was based on the implicit assumption that the alcohol variable(s) of interest were exogenous right-hand side variables. Using "disorder" variables like alcoholism measures as regressors (e.g., Benham and Benham 1982, Mullahy and Sindelar 1991*b*) is sometimes justified as an extension of the tradition of including measures of health capital as supposedly exogenous regressors in earnings models, evident at least as early as the Bartel and Taubman (1979) study discussed earlier. The inclusion of alcohol consumption measures in labor market success equations is perhaps best grounded from the theoretical perspective of a conditional labor supply model (see Browning and Meghir 1991 and Deaton and Muellbauer 1980 for further discussion).

Yet, whether alcoholism or alcohol consumption is the drinking measure of interest, the prospect that the alcohol variable is correlated with unobserved determinants of the labor market outcome of interest looms large. In the health capital literature, the possibility of reverse or simultaneous causation is obvious: Health problems may both determine and be determined by success in the labor market. When alcohol consumption is of interest, standard consumer theory sug-

gests that it along with other consumption and labor supply decisions are simultaneously determined as functions of prices and other constraints.

In both scenarios, maintaining econometric exogeneity of the alcohol measure is, therefore, tenuous. Since none of the studies surveyed above was based on experimental data, and since there has been little use to date of instrumental variable methods as controls for possible endogeneity,[28] it is useful to exposit the basic elements of "simultaneous equations bias" or, more descriptively, regression with unobservables correlated with observables.

Suppose the true structural model is given by

$$y = \alpha A + \beta X + \varepsilon$$

where y is some measure of labor market success (wage rate, earnings, etc.), A is some measure of alcohol use of interest (e.g., alcohol consumption or a binary indicator of alcoholism), X is a scalar regressor, ε is the regression error, and α and β are unknown parameters to be estimated. To make the discussion concrete, suppose $\alpha < 0$. If ordinary least squares (OLS) is used in estimation, the point estimate of α is

$$\begin{aligned}\hat{\alpha} &= [\,\alpha(m_{XX}m_{AA} - m_{AX}^2) + m_{XX}m_{A\varepsilon} \\ &\quad - m_{AX}m_{X\varepsilon}\,] / [\,m_{XX}m_{AA} - m_{AX}^2\,] \\ &= \alpha + [\,(m_{XX}m_{A\varepsilon} - m_{AX}m_{X\varepsilon}) / \\ &\quad (m_{XX}m_{AA} - m_{AX}^2)\,]\end{aligned}$$

where $m_{RS} = \sum_{i=1}^{N} R_i S_i$. Let $\sigma_{ij} = plim(m_{ij}/N)$ and, without loss of generality, define units of X such that $\sigma_{XX} = 1$. The large sample central tendency of this estimator is given by its *plim* as

$$\begin{aligned}plim\,(\hat{\alpha}) = \alpha + [\,&(\sigma_{A\varepsilon} - \sigma_{AX}\sigma_{X\varepsilon}) \\ &\div (\sigma_{AA} - \sigma_{AX}^2)\,]\end{aligned}$$

The denominator of the bracketed term is the determinant of a covariance matrix and is therefore necessarily positive. Thus, the sign of the OLS inconsistency (i.e., $plim(\hat{\alpha}) - \alpha$), in the case where A is correlated with the unobservable ε, is determined by the sign of the numerator, $\sigma_{A\varepsilon} - \sigma_{AX}\sigma_{X\varepsilon}$.

The simplest case to consider is where A and X are uncorrelated so that $\sigma_{AX} = 0$. In this instance, the sign of the inconsistency is given by the sign of $\sigma_{A\varepsilon}$. If A is positively correlated with unobservables ($\sigma_{X\varepsilon} > 0$), then the inconsistency is positive and the sign of $plim(\hat{\alpha})$ is indeterminate since $\alpha < 0$ by assumption; if A is negatively correlated with unobservables ($\sigma_{X\varepsilon} < 0$), then the inconsistency is negative and $plim(\hat{\alpha}) < 0$. A priori, correlations of either sign are possible, of course.

Consider the intuition underlying and a fable consistent with each. When $\alpha < 0$ and $\sigma_{A\varepsilon} > 0$, then A—which independently has negative impacts on y—is positively correlated with an unobserved variable that independently exerts a positive influence on y. An example of this might be that ε is "drive to succeed," that

[28] *"Endogeneity" most fundamentally describes a situation where one or more measured regressors is correlated with unobservables, i.e., with the equation error.*

individuals driven to succeed also tend to drink more on average than those not driven to succeed, and that with a ceteris paribus reduction in drinking these individuals would have even greater earnings. Since "drive" is not measured in the data, the independent negative contribution of A ($\alpha < 0$) is obscured in estimation by the positive correlation between alcohol consumption and "drive." As indicated above, the resulting point estimate will at a minimum have a central tendency less negative than α itself, with a positive central tendency of $\hat{\alpha}$ perfectly possible if the correlation between A and ε is sufficiently strong.

When $\alpha < 0$ and $\sigma_{A\varepsilon} < 0$, then A is negatively correlated with an unobserved variable that independently exerts a positive influence on y. A tale consistent with this structure is that ε again is "drive to succeed," but now individuals driven to succeed tend to have relatively low values of A. In this case, the independent negative contribution of A ($\alpha < 0$) is reinforced in estimation by the negative correlation between alcohol consumption and "drive." The resulting inconsistent point estimate will suggest an even greater negative effect of A on y than is truly the case.

Although demonstrated in a relatively unstructured framework, this simple but obvious case of underidentification indicates how the key biases depend essentially on basic correlations like $\sigma_{A\varepsilon}$. Illustrated above in the context of a simple single equation earnings model, the implications may be even more lucid in an explicitly simultaneous framework.[29]

A fitting example of such a model is one that attempts to test whether problem drinking (D) is a structural determinant of unemployment (U), unemployment is a structural determinant of problem drinking, both, or neither. That is, is unemployment a "stressor" that brings on drinking problems, or does problem drinking result in substandard workplace performance that in turn results in unemployment? As suggested earlier, the potential for reverse causality seems particularly great in models where employment outcomes are regressed on alcohol variables, and a simple two-equation model underscores precisely how such econometric problems might arise.

In a linear model context, the main objective in such an exercise might be to identify and consistently estimate γ_1 and δ_1 in the two-equation structural model

$$U = \gamma_1 D + \gamma_2 X + \eta_U$$

and

[29] *The most obvious "simultaneity" is, of course, that the demand for alcoholic beverages will depend on consumer or household income, since the latter is a determinant of the budget set (see Leung and Phelps in this volume for a survey of the literature on the demand for alcoholic beverages). Variations on such a model would allow for the possibility that income not only defines the budget set but also affects choice by creating increasing levels of "stress" as income decreases. Stress, which is undesirable, can be temporarily alleviated by consumption of alcoholic beverages, which may also be demanded for other reasons as well. Under some reasonable assumptions, the income elasticity of demand (at given levels of income and alcohol consumption) will be lower in such a model than in a model that allows only for the budget constraint effect.*

$$D = \delta_1 U + \delta_2 X + \eta_D$$

where X is a vector of covariates and η_U and η_D are unobservables.[30] Unless $\sigma_{\eta_U \eta_D} = 0$, D is correlated with η_U (and U is correlated with η_D) so that the previous comments on regressors correlated with unobservables apply. The fundamental econometric problem is that without a priori restrictions on γ_2, δ_2, or $\sigma_{\eta_U \eta_D}$, this structural model is underidentified.

There are, at least in principle, several possible solutions to the underidentification that arises when regressors are theoretically or behaviorally likely to be correlated with unobservables. The strategy chosen in practice, of course, will be dictated by the nature of the available data, but it is worth summarizing here the basic options that might be available to researchers (see Mullahy and Manning in press for further discussion).

Experimental data that by design force orthogonality between regressors and unobservables are an ideal solution, but one rarely available in cases of interest to those analyzing problems involving alcohol and the labor market. Natural experiments offer a second prospect for identification, but to date none has been identified as promising in resolving such issues.[31] A third strategy is to utilize longitudinal data. If that part of the unobservables correlated with the regressors has a time-invariant or fixed-effect character, then by using longitudinal data one may—depending on how the fixed-effect enters the model—be able to "wash out" the fixed effects by differencing or like methods.

Instrumental variable estimation, which essentially combines the idea of natural experiments with a priori restrictions on γ_2 or δ_2, is yet another strategy that can be used in this context. If it is reasonable, for instance, to maintain that some regressor or regressors in X are structural determinants of U but not of D—i.e., that some element(s) of δ_2 are zero on a priori grounds—then using these "excluded" regressors as instruments for U in the D model allows consistent estimation of δ_1 and δ_2 if the a priori restrictions are valid. Rarely does economic theory afford clear guidance as to the choice of appropriate zero restrictions, so strategies used in practice will most times smack of being ad hoc.

---

[30] *More generally, the objective is to identify and estimate the conditional distributions Pr(U|D,X) and Pr(D|U,X) or properties thereof, like expectations. The linear models are assumed only for expositional simplicity.*

[31] *While one is hard-pressed to imagine a situation where a natural experiment would randomly assign individuals to alcohol abuse regimes in a manner independent of their employment status, it is more tenable to conceive of situations where individuals might be "naturally" assigned to employment or unemployment in a manner independent of their problem drinking. For instance, a large company that lays off a sizable chunk of its workforce may be bound by union rules to distribute pink slips in a manner basically unrelated to drinking problems. Laid off workers would then be the "treatment" group while remaining workers could serve as controls. In statistical terms, layoffs of this sort would effectively orthogonalize U and* η_D *in the problem drinking equation and thus offer prospects for identifying key parameters of that equation.*

Nonetheless, certain restrictions are probably "more reasonable" than others and, at a minimum, two competing sets of estimates (one using instrumental variables and one not) provide the analyst with the basis of possibly useful specification tests.[32]

Issues in functional form specification

Another econometric issue that has clear relevance in the empirical literature surveyed here concerns the specification of the functional form of a probability model's corresponding regression or conditional mean function. Consider the following simple example. Suppose the individual's wage rate (w) is a simple quadratic function of alcohol consumption (A),

$$w = \alpha + \beta A + \gamma A^2 + \varepsilon$$

and, for this example, assume $\sigma_{A\varepsilon} = 0$. Suppose the true model is $\alpha = 6$, $\beta = 16$, $\gamma = -1$, and $\varepsilon \equiv 0$ (i.e., there are no unobservables in the data), so that w increases with A up to $A = 8$ and declines thereafter. If the bulk of the population drinks less than eight drinks, then it is quite easy to miss statistically the negative relationship between w and A that ultimately occurs for "heavy" drinkers if the correct functional form is not used.

As a concrete example, suppose in a sample of seven individuals the observations on (A,w) are (0,6), (2,34), (4,54), (6,66), (8,70), (10,66), and (12,54). Four observations are on (w,A) on the upward-sloping part of the wage profile, one on (w,A) at the peak, and two on (w,A) on the downward-sloping segment. Least-squares regression of w on A, A^2, and a constant term will recover exactly the population parameters ($\alpha = 6$, $\beta = 16$, $\gamma = -1$), since the data in this example are errorless. However, if the wrong functional form is assumed and w is regressed only on A and a constant term (i.e., $\hat{\gamma}$ is restricted to be zero), then the estimated parameters are $\hat{\alpha} = 26$, $\hat{\beta} = 4$, suggesting a positive relationship between w and A.

This simple example should serve to show how inferences about relationships between alcohol use and labor market success may depend critically on some fairly subtle statistical issues like functional form specification. One striking example of the possible importance of such functional form issues is Cook's (1991) study in which a nonlinear relationship between alcohol use and income was apparent in the simple graphical analysis. Given sufficiently large sample sizes, the obvious strategy in resolving this functional form issue is to utilize nonparametric or semi-nonparametric regression methods (see Hardle 1990 and Hastie and Tibshirani 1990).

Model heterogeneity

The last econometric issue to be addressed here concerns the possibility and implications of model heterogeneity. This issue

[32] *For instance, in a model akin to that described here, an analyst might consider restricting the coefficient on the State unemployment rate to be zero in an alcohol abuse equation and restricting the coefficients on State beer taxes and parental problem drinking measures to be zero in an employment status equation. This is the approach being pursued by Jody Sindelar and me in ongoing research using the 1988 NHIS.*

can be described from several perspectives, perhaps the most popular being that of a stochastic parameters model. Another perspective that is at least as useful, however, derives from the literature on conditional quantile estimation (see Hendricks and Koenker 1992, Koenker and Bassett 1982, and Powell 1991), whose key insight for purposes at hand is as follows. One fairly subtle implication of a standard form of heteroskedasticity is that it induces changes in the probability distribution on the partial responses of the dependent variable in response to changes in an independent variable; that is, there are heterogeneous responses. To fix ideas, suppose a model of earnings is given by

$$y = \alpha A + \zeta Z + (\xi_1 A + \xi_2 Z)\varepsilon$$
$$= X\beta + (X\xi)\varepsilon$$

where A is an alcohol measure, Z is a vector of other covariates, $X = [A,Z]$, and, for purposes of this example, $E(\varepsilon | X) = 0$.

Define $\int_{(-\infty, Q_q(\varepsilon))} dF(\varepsilon) = q$, so that $Q_q(\varepsilon)$ is the q-th quantile of the probability distribution of ε, $F(\varepsilon)$. It then follows that the q-th conditional quantile of y given X is

$$Q_q(y|X) = X\beta + (X\xi)Q_q(\varepsilon)$$

Usually economists focus on estimating parameters such as

$$\partial E(y|X)/\partial X_j = \beta_j$$

but consider more generally the partial derivative

$$\partial y/\partial X_j = \beta_j + \xi_j \varepsilon$$

In this context, it is reasonable to speak of $\partial y/\partial X_j$ as itself having a probability distribution whose quantiles are given by

$$Q_q(\partial y/\partial X_j) = \beta_j + \xi_j Q_q(\varepsilon) \equiv b_{qj}$$

The interpretation of the b_{qj} is that it summarizes the partial response between y and X_j at the q-th quantile of the distribution of the unobservables ε. It is these "coefficients" that are estimated by the regression quantile estimators, $\hat{b}_{qj}$, that were introduced to economists by Koenker and Bassett (1982).

Why might this approach be of interest to analysts evaluating relationships between alcohol use and the labor market? It is commonplace in the alcohol literature to encounter statements that more or less convey the message that alcohol affects different people in different ways. Casual empiricism certainly suggests that the subpopulation of alcoholics comprises economic successes as well as economic failures. Accordingly, attempts to characterize relationships between problem drinking and economic success using econometric models that do not admit parameter heterogeneity may result in serious model misspecification.[33]

[33] *Applications of quantile estimators or other varying parameter models to the study of these relationships have only recently begun; see, for example, Grossberg and Mullahy (1992).*

CONCLUSIONS AND DIRECTIONS FOR FUTURE RESEARCH

This paper has attempted to provide a broad survey of what is understood to date about the relationships between drinking and labor market outcomes. Given what may be at stake, the empirical evidence on such relationships is quite modest and, as suggested in the discussion in the second section of this chapter, such evidence as presently exists is far from conclusive. Perhaps the most optimistic conclusion that can be offered is that considerable progress has been and is being made in understanding these important relationships. The fact that the lion's share of the empirical work in this field is less than a decade old augurs well for what we might expect to be known 10 years hence if future research and data-gathering efforts continue at a healthy pace.

The existing research surveyed earlier and the conceptual framework offered above indicate several directions in which future research efforts might be fruitfully focused. To close this essay, at least five such directions are worth sketching.

The first is data. As suggested above, many of the seemingly disparate results appearing in the literature owe simply to the fact that there have been very few commonalities among the various datasets that have been used to date in terms of their labor market outcome measures, their alcohol use measures, or both. Moreover, with the exception of the NLSY data used by Cook and Moore (and, to a minor degree, the ECA data as well), reliable longitudinal data have not been readily available. Absent longitudinal data, many interesting hypotheses about alcohol and the labor market [e.g., the direction(s) of the employment/alcohol abuse relationship] would at a minimum require econometrically valid instrumental variables in the data, but these too have for the most part been lacking. Any future data-gathering efforts should place high priority on retaining detailed geographical information on prices, wages, local labor market conditions, and other possible (quasi-) natural experiments that might be exploited for use as instruments. It may be that the National Longitudinal Alcohol Epidemiologic Survey (NLAES), now under way, will offer researchers a level of data sophistication well beyond that available in existing data sources.

The second issue is also a data issue of sorts, and specifically concerns the fact that empirical studies to date have focused largely—albeit not exclusively—on samples of males. Given that most work has focused on income and earnings, this emphasis is perhaps understandable. Studying the relationships between problem drinking and labor market success for females is more difficult than for males for at least two reasons: Females' labor supply behavior is more complex to analyze than that of males, and alcohol problems are relatively less common among the female population.[34] While some initial attempts at gender comparisons have

[34]*See Grant et al. (1991) who used data from the 1988 NHIS as the basis of their estimates that national DSM-III(R) 1-year prevalences of total alcohol abuse and/or dependence are 13.4 percent and 4.4 percent for males and females, respectively.*

appeared in the literature (e.g., Mullahy and Sindelar 1991*a*), much remains to be done in this area of inquiry.

The third area in which further research efforts should be fruitful is labor supply and its relationships to alcohol use. As suggested above, some results are now available on basic issues like market participation and employment, but there is no known empirical evidence on how hours and weeks of labor supply relate to alcohol use or alcoholism. How alcohol use affects the variability of employment over a given time period is also an important issue to address from a productivity perspective, but apart from the evidence presented above on alcohol use and absenteeism, the information base here is quite sparse as well. Finally, as suggested earlier, it would be extremely interesting to have reliable estimates of the labor supply decisions and alcohol use patterns of all adults in the household; most existing studies are constrained by having information on only one adult's alcohol use and/or labor supply patterns in the data.

A fourth consideration for researchers involves the conceptualization of alcohol use, alcohol consumption, alcoholism, etc., for use in data analysis. Some of the disparate empirical results in the literature are surely due to the fact that vastly different measures of "the alcohol variable" have been included in regression models (see the paper by Sindelar in this volume for discussion and an example). Tremendous progress has been made in designing survey instruments to provide "diagnoses" of alcohol-related problems, and the data sets containing such measures that will ultimately be available (e.g., NLAES) should be of enormous utility to researchers.[35] In the meantime, however, distinctions between and among drinking, light drinking, heavy drinking, frequency of drinking, alcoholism, alcohol dependence, and alcohol abuse are crucially important to draw in undertaking and in interpreting empirical work.[36]

The fifth dimension—whose surface has only slightly been scratched thus far—is human capital accumulation and how it relates to alcohol use. As suggested above, the many tough conceptual problems involved here are only compounded by data constraints. However, if the arguments by Cook and Moore (1992) and Mullahy and Sindelar (1990*a*,*b*) are on the mark, then inquiry into the manner in which one's own, one's parents', or one's peers' alcohol use might affect the quality and quantity of schooling attainment, the formation of households, the level and quality of labor market experience, and other key components of human capital is of the utmost importance in determining the productivity effects of problem drinking. Indeed, after more evidence is in, it would not be surprising to this analyst to find that diminished human capital is the

[35]*For example, DSM-III(R)-based and ICD-10-based instruments are now in use; see Grant and Towle (1990) for discussion.*

[36]*See Heather and Robertson (1989) for an excellent and balanced recent discussion. At this point it is also worth noting the recent economic literature on addiction and rational addiction; see Becker and Murphy (1988) and Grossman (chapter 4).*

single most important consequence of problem drinking. If so, then the economic merits of prevention policies become all the more obvious.

ACKNOWLEDGMENTS

This research has been supported in part by a university fellows grant from Resources for the Future, and draws on joint research with Jody Sindelar funded in part by grant R01AA08394 from the National Institute on Alcohol Abuse and Alcoholism to Yale University. I am indebted to Greg Bloss, Phil Cook, Dale Heien, Mike Hilton, Don Kenkel, Jack Mullahy, and Jody Sindelar for helpful comments and suggestions as the paper evolved, and to Randy Ellis and Gail Mitchell Hoyt for their help in tracking down some of the literature reviewed here.

REFERENCES

Aaron, P., and Musto, D. Temperance and prohibition in America: A historical overview. In: Moore, M.H., and Gerstein, D.R., eds. *Alcohol and Public Policy: Beyond the Shadow of Prohibition.* Washington, DC: National Academy Press, 1981. pp. 127–181.

Akerlof, G.A., and Yellin, J.L. *Efficiency Wage Models of the Labor Market.* Cambridge: Cambridge University Press, 1986.

Andrews, D.W.K. Power in econometric applications. *Econometrica* 57(5):1059–1090, 1989.

Bartel, A., and Taubman, P. Health and labor market success: The role of various diseases. *Rev Econ Stat* 61(1):1–8, 1979.

Bartel, A., and Taubman, P. Some economic and demographic consequences of mental illness. *J Labor Econ* 4(2):243–256, 1986.

Becker, G.S. *Human Capital.* 2d ed. Chicago: University of Chicago Press, 1975.

Becker, G.S., and Murphy, K.M. A theory of rational addiction. *J Polit Econ* 96(4):675–700, 1988.

Benham, L., and Benham, A. Employment, earnings, and psychiatric diagnosis. In: Fuchs, V., ed. *Economic Aspects of Health.* Chicago: University of Chicago Press, 1982. pp. 203–220.

Berger, M.C., and Fleisher, B.M. Husband's health and wife's labor supply. *J Health Econ* 3(1):63–75, 1984.

Berger, M.C., and Leigh, J.P. The effect of alcohol use on wages. *Appl Econ* 20(10):1343–1351, 1988.

Berkowitz, M., and Johnson, W.G. Health and labor force participation. *J Hum Resources* 9(1):117–128, 1974.

Berry, R., and Boland, J. *The Economic Costs of Alcohol Abuse.* New York: Free Press, 1977.

Browning, M., and Meghir, C. The effects of male and female labor supply on commodity demands. *Econometrica* 59(4):925–951, 1991.

Cook, P.J. The social costs of drinking. In: *Expert Meeting on Negative Social Consequences of Alcohol Use.* Oslo: Norwegian Ministry of Health and Social Affairs, 1991. pp. 49–81.

Cook, P.J., and Moore, M.J. "Drinking and Schooling." Presented at the Third Annual Health Economics Workshop, Johns Hopkins University, May 1992.

Deaton, A., and Muellbauer, J. *Economics and Consumer Behavior.* Cambridge: Cambridge University Press, 1980.

Gill, A.M., and Michaels, R.J. Does drug use lower wages? *Ind Labor Relations Rev* 45(3):419–434, 1992.

Glazer, J., and Weiss, A. "A Welfare Analysis of Drug Testing in the Workplace." Mimeo, Boston University, 1992.

Grant, B.F.; Harford, T.C.; Chou, P.; Pickering, R.; Dawson, D.A.; Stinson, F.S.; and Noble, J. Prevalence of DSM-III-R alcohol abuse and dependence: United States, 1988. *Alcohol Health Res World* 15(1):91–96, 1991.

Grant, B.F., and Towle, L.H. Standardized diagnostic interviews for alcohol research. *Alcohol Health Res World* 14(4):340–347, 1990.

Grossberg, A., and Mullahy, J. "Analyzing the Determinants of Earnings via Conditional Quantiles." Mimeo, Trinity College, 1992.

Grossman, M. On the concept of health capital and the demand for health. *J Polit Econ* 80(2):223–255, 1972.

Hardle, W. *Applied Nonparametric Regression.* Cambridge: Cambridge University Press, 1990.

Harwood, H.J.; Cruze, A.M.; Kristiansen, P.L.; Collins, J.J.; and Jones, D.C. *Economic Costs to Society of Alcohol and Drug Abuse and Mental Illness: 1980.* Research Park, NC: Research Triangle Institute, 1984.

Hastie, T.J., and Tibshirani, R.J. *Generalized Additive Models.* London: Chapman and Hall, 1990.

Heather, N., and Robertson, I. *Problem Drinking.* 2d ed. Oxford: Oxford University Press, 1989.

Heien, D.M., and Pittman, D.J. The economic costs of alcohol abuse: An assessment of current methods and estimates. *J Stud Alcohol* 50(6):567–579, 1989.

Hendricks, W., and Koenker, R. Hierarchical spline models for conditional quantiles and the demand for electricity. *J Am Stat Assoc* 87(417):58–68, 1992.

Hingson, R.W.; Lederman, R.I.; and Walsh, D.C. Employee drinking patterns and accidental injury: A study of four New England States. *J Stud Alcohol* 46(4):298–303, 1985.

Hoyt, G.M. "The Worker, the Firm, and the Decision to Use Drugs." Ph.D. dissertation, Department of Economics, University of Kentucky, 1992.

Kaestner, R. The effect of illicit drug use on the wages of young adults. *J Labor Econ* 9(4):381–412, 1991.

Killingsworth, M.R. *Labor Supply.* Cambridge: Cambridge University Press, 1983.

Koenker, R., and Bassett, G. Robust tests for heteroskedasticity based on regression quantiles. *Econometrica* 50(1):43–61, 1982.

Luft, H. The impact of poor health on earnings. *Rev Econ Stat* 57(1):43–47, 1975.

Manning, W.G.; Keeler, E.B.; Newhouse, J.P.; Sloss, E.M.; and Wasserman, J. *The Costs of Poor Health Habits.* Cambridge, MA: Harvard University Press, 1991.

Mincer, J. *Schooling, Experience, and Earnings.* New York: Columbia University Press for NBER, 1974.

Model, K. "Drug and Alcohol Use and Income: A Review of the Literature and Some New Estimates." Chapter III, Ph.D. dissertation, Harvard University, 1992.

Mullahy, J., and Manning, W.G. Statistical issues in cost-effectiveness analyses. In: Sloan, F., ed. *Issues in Cost-Effectiveness Analysis* (tentative title), in press.

Mullahy, J., and Sindelar, J. Lifecycle effects of alcoholism on education, earnings, and occupation. *Inquiry* 26(2):272–282, 1989.

Mullahy, J., and Sindelar, J. An ounce of prevention: Productive remedies for alcoholism. *J Policy Analysis Manage* 9(2):249–253, 1990*a*.

Mullahy, J., and Sindelar, J. Direct and Indirect Effects of Alcoholism on Human Capital. Submitted for publication. 1990*b*.

Mullahy, J., and Sindelar, J. Gender differences in labor market effects of alcoholism. *Am Econ Rev* 81(2):161–165, 1991*a*.

Mullahy, J., and Sindelar, J. Alcoholism, work, and income. *J Labor Econ*, 1991*b*.

Mullahy, J., and Sindelar, J. "Health Capital, Risk Aversion, and the Variance of Income: Assessing Some Welfare Costs of Alcoholism and Poor Health." 1992.

Mushkin, S.J. Health as an investment. *J Polit Econ* 70(5):129–157, 1962.

Parsons, D.O. Health, family structure, and labor supply. *Am Econ Rev* 67(4):703–712, 1977.

Pencavel, J. Labor supply of men: A survey. In: Ashenfelter, O., and Layard, R., eds. *Handbook of Labor Economics.* Vol. 1. Amsterdam: North-Holland, 1986. pp. 3–102.

Piacentini, J.S., and Cerino, T.J. *EBRI Databook on Employee Benefits.* Washington, DC: Employee Benefit Research Institute, 1990.

Plant, M.A. Alcoholism and occupation: A review. *Br J Addict* 72(2):309–316, 1977.

Powell, J. Estimation of monotonic regression models under quantile restrictions. In: Barnett, W.A., et al., eds. *Nonparametric and Semiparametric Methods in Econometrics and Statistics.* Cambridge: Cambridge University Press, 1991. pp. 357–384.

Register, C.A., and Williams, D.R. Labor market effects of marijuana and cocaine use among young men. *Ind Labor Relations Rev* 45(3):435–448, 1992.

Rice, D.P.; Kelman, S.; Miller, L.S.; and Dunmeyer, S. *The Economic Costs of Alcohol and Drug Abuse and Mental Illness: 1985.* DHHS Pub. No. (ADM)90–1694. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1990.

Rubin, D.B. Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol* 66(5):688–701, 1974.

Russell, L.B. *Is Prevention Better Than Cure?* Washington, DC: Brookings Institution, 1986.

Smith, J.P., ed. *Female Labor Supply: Theory and Estimation.* Princeton: Princeton University Press, 1980.

Strauss, J. Does better nutrition raise farm productivity? *J Polit Econ* 94(2):297–320, 1986.

U.S. Department of Health and Human Services. *Sixth Special Report to the U.S. Congress on Alcohol and Health.* DHHS Pub. No. (ADM)87–1519. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1987.

U.S. Department of Health and Human Services. *Seventh Special Report to the U.S. Congress on Alcohol and Health.* DHHS Pub. No. (ADM)90–1656. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1990.

Willis, R.J. Wage determinants: A survey and reinterpretation of human capital earnings functions. In: Ashenfelter, O., and Layard, R., eds. *Handbook of Labor Economics.* Vol. 1. Amsterdam: North-Holland, 1986. pp. 525–602.

7

Minimum Drinking Age and Alcohol Availability to Youth: Issues and Research Needs

Alexander C. Wagenaar, Ph.D.

INTRODUCTION

Most youth in the United States consume beverage alcohol, and the prevalence of use increases rapidly with age. Approximately 6 percent of 10- to 11-year-olds are current users (use in the past 30 days) of alcohol; the rate increases to about 25 percent at ages 12 to 14, and to 55 percent at ages 15 to 17 (Dryfoos 1990). Among high school seniors, 90 percent report consuming alcohol at some point in their lives, and 57 percent report being current drinkers (Johnston et al. 1991). Moreover, 35 percent report regularly becoming intoxicated (five or more drinks per occasion). Junior and senior high-school-age youths drink an estimated 31.2 million gallons of wine coolers annually (35 percent of all wine coolers sold) and 1.1 billion cans of beer annually (2 percent of all beer sold; Kusserow 1991). As a result of high rates of drinking, youth experience high rates of health and social problems associated with alcohol. Motor vehicle crashes are the leading cause of death for teenagers (Baker et al. 1984), with one-third to one-half involving alcohol [National Highway Traffic Safety Administration (NHTSA) 1985]. Other leading causes of death and long-term disability for youth, such as suicides, homicides, assaults, drownings, and recreational injuries, frequently involve alcohol (US DHHS 1990). Although exact quantification of the health burden of underage alcohol consumption is difficult, underage alcohol use results in substantial premature deaths, disability, preventable hospitalizations, fiscal costs, and human suffering.

However, trauma morbidity and mortality among underage drinkers is only one of many reasons for concern. There may be deleterious effects of youth alcohol use on essential dimensions of physical, cognitive, and social development (Semlitz and Gold 1986). There may also be longer-term effects on risk of infectious diseases, including acquired immunodeficiency syndrome (AIDS; Stall et al. 1986); chronic diseases, such as chronic liver disease and cirrhosis (Grant et al. 1988); and cancers of the nasopharynx, larynx, liver, and esophagus (Driver and Swann 1987). Although there remains a diversity of opinion, extant lit-

erature provides considerable support for the stage drug use theory—that youths systematically move through a series of stages of drug use, beginning with tobacco and alcohol before moving on to other illegal drugs, such as marijuana, cocaine, and heroin (Yamaguchi and Kandel 1984; Kandel 1988). Few would accept a "gateway theory" that early tobacco or alcohol use inevitably causes one to move on to use of other drugs; however, early use of drugs that are legal for adults does increase the risk of later use of illegal drugs. Thus, preventing early drinking may also prevent use of a variety of other drugs with serious health and social consequences.

There are indications that early drinkers are at higher risk of subsequent addiction to, or problems with, alcohol (Robins 1978; Gonzalez 1989). Delaying the age of initiation into regular drinking may, therefore, have beneficial effects in reducing rates of addiction to alcohol. It is not only age of first drink that may be important. Recent research has shown that age of first easy access to alcohol is related to later rates of drinking. O'Malley and Wagenaar (1991) found that early legal access to alcohol (at age 18) was associated with higher rates of drinking later (at ages 21 to 25). Youth who did not experience legal access to alcohol until age 21 (as a result of external policy changes raising the legal age) not only reduced their drinking during ages 18 to 20, but exhibited lower rates of drinking during the ages 21 to 25, after they had attained legal age. In short, in addition to the "age of initiation effect," there may be a further "age of legal initiation effect." The implication is that even if first use of alcohol occurs at an early age, preventing the emergence of a pattern of regular drinking during the teenage years may have further beneficial effects during adulthood, in addition to the teenage injuries and other problems that may be averted.

It is worth noting that the recent findings of O'Malley and Wagenaar (1991) do not support two alternative hypotheses regarding the effects of raising the drinking age. The first is a "rubber band theory": Prohibiting youth from drinking until age 21 will result in a pent-up demand for the "forbidden fruit." At age 21, therefore, the pent-up pressure to drink would spring loose, with 21-year-olds drinking at significantly higher rates than they would have if given access to alcohol earlier. This theory is not supported by the evidence; in fact, drinking at age 21 is lower when the legal age is 21 than when the legal age is below 21. The second hypothesis is similar yet conceptually distinct: Raising the legal age to 21 simply delays until age 21 the alcohol problems that would have occurred during ages 18 to 20. The contention is not that the higher age makes things worse, only that it has no net effect except shifting the time at which alcohol problems are manifested. Again, the evidence does not support this hypothesis. Traffic crashes and other alcohol problems averted by the "age 21" policy during the period when persons are aged 18 to 20 do not accumulate simply to emerge later, but are permanently averted.

The importance of accessibility of alcohol is shown by the results of a large number of studies of the effects of lowering and subsequently raising the legal age for purchasing and consuming alcohol, and by the findings of followup studies showing that youth access remains high, with little enforcement of the drinking age. Drinking rates and frequency of alcohol-involved problems (particularly traffic crashes) among 18- to 20-year-olds increased after the legal age was lowered in 29 States in the early 1970's. Policy changes in the late 1970's and early to mid-1980's returned the legal age to 21 in all States. Studies prior to the mid-1980's were reviewed in detail in my 1983 book (Wagenaar 1983) and are included here for completeness. In addition, studies of the raised minimum drinking age published since the early 1980's are summarized.

EFFECTS OF MINIMUM DRINKING AGE ON ALCOHOL USE

Implicit in research on the effects of the drinking age on alcohol-related problems among youth is the assumption that the legal age influences the amount or the pattern of alcohol consumption by young people. Studies of the impact of the drinking age on alcohol consumption among youth are less methodologically sophisticated than is research that assesses the impact of drinking-age changes on traffic crashes. The main reason for this difference is that age-specific data on longitudinal trends in alcohol consumption have only recently become available.

This review begins with a discussion of studies undertaken in the 1970's, most of which analyzed reductions in the drinking age. This is followed by a discussion of studies conducted in the 1980's and early 1990's, most of which analyzed increases in the minimum age. The earlier group of studies (1970's) can be further subdivided into those that were based on survey data and those that were based on aggregate statistics. The latter group of studies (1980's and 1990's) includes studies of aggregate statistics and surveys, with the survey studies further divided into those that surveyed the general youth population and those that surveyed college student populations.

Surveys in the 1970's of the effects of lowering the drinking age on alcohol consumption among youth produced conflicting results. Wolfe and Chapman (1973) surveyed Michigan high school students in 1971, before the drinking age was lowered, and again in 1973, after the reduction in drinking age. A substantial increase in the frequency of drinking and an increase in the quantity of alcohol consumed per occasion were found after the legal age was lowered. The authors pointed out, however, that the increases were consistent with the preexisting trends in alcohol use among youth and therefore could not be unambiguously attributed to the lowered legal drinking age.

Smart and Schmidt (1975) conducted a similar before-and-after survey of Toronto junior and senior high school students. After a reduction in the drinking age, 40 percent of the students reported no change in drinking patterns, 20 percent reported drinking more, 4 percent reported drinking less, and 9 percent indicated

that they had started drinking since the drinking age had been reduced (the other 27 percent were abstainers). Smart and Schmidt also surveyed college students, the majority of whom reported no change in frequency or quantity of alcohol consumption, although 55 percent did report increased patronization of public drinking establishments since the legal change.

McFadden and Wechsler (1979) surveyed Massachusetts teenagers in 1965, 1970, and 1974. Alcohol consumption for this group increased between 1965 and 1970, when there was no change in the drinking age, as well as between 1970 and 1974, when the legal age was reduced from 21 to 18. Those authors also surveyed New England college students in 1977 and found that students from States with a low legal drinking age consumed alcohol more frequently than students from States with a high drinking age.

Rooney and Schwartz (1977) surveyed high school students in five selected States: three with minimum legal drinking ages at 18, one with a minimum age at 20, and one with a minimum age at 21. The analyses rested on cross-sectional comparisons between data from the first three States (regarded as low drinking age States) and the latter two States (regarded as high drinking age States). Findings ran contrary to the thesis that a higher drinking age would be associated with lower consumption and fewer problems. In the higher drinking age States, weekly alcohol consumption was slightly more common (reported by 47 percent versus 42 percent of students), the prevalence of drinking was higher (reported by 81 percent versus 76 percent of students), and the incidence of alcohol-related problems was higher than in the lower age States. The authors concluded that a high drinking age did not have a beneficial effect in controlling alcohol consumption among young people and that it might even have adverse effects.

Results more favorable to the thesis that higher drinking ages would covary with preventive effects were obtained by Maisto and Rachal (1980) in their analyses of a probability sample representative of high school students in the United States. They found that students in States with a higher legal drinking age were more likely to be abstainers, were less likely to be heavy drinkers, and experienced intoxication less frequently than students in States with lower drinking ages. The authors concluded that legal availability of beverage alcohol, as reflected in the drinking age, is associated with the drinking practices of young people.

Vingilis and Smart (1981) surveyed junior and senior high school students in Ontario in 1977, when the drinking age was 18, and again in 1979, just 2 months after the legal age had been increased to 19. The number of drinkers aged 18 and 19 and the proportion who reported getting "tight" in the previous month decreased significantly after the higher age limit was implemented. In contrast, the number of drinkers aged 17 and under and the proportion in this group who reported getting "tight" increased significantly from 1977 to 1979. Three considerations should be kept in mind when interpreting the results of this study. First, students were questioned just 2 months

after the drinking age change; results therefore only apply to the short-term impact of the new law. Second, 18-year-olds, the only age category directly affected by the drinking age increase from 18 to 19, were not analyzed separately, but were combined with 19-year-olds in the analyses. Third, Ontario's new drinking age law included a grandfather clause; as a result, few 18-year-olds actually were barred from drinking in the first 2 months after the law took effect.

Hingson et al. (1983) surveyed Massachusetts youth before and after an increase in the legal drinking age. No significant changes in reported alcohol consumption among 16- to 19-year-olds (analyzed as a single group) occurred after the drinking age was raised from 18 to 20. Separate results for the focal group of interest, 18- and 19-year-olds, were not presented.

Survey-based research from the 1970's has given a mixed picture of findings. Findings indicate that instances of lowering the drinking age in the early 1970's were accompanied by increased youth drinking, but upward trends in alcohol consumption levels that preceded these legal changes make it difficult to attribute unambiguously the postintervention findings to the effect of legal change. Studies employing cross-sectional analyses are methodologically weaker than pre-post designs and have shown that increased minimum age both does and does not covary with decreased youth drinking. Pre-post surveys of the effects of raising drinking ages in the later 1970's have also been mixed, showing no effect in one instance and a preventive effect in another.

The second major type of data used in the 1970's to assess the impact of reduced drinking ages on alcohol consumption patterns is aggregate sales volumes. Smart and Schmidt (1975) compared Ontario beverage alcohol shipments before and after a reduction in the legal age. Apparent consumption during the first 5 months after the legal change was higher than expected on the basis of the prechange figures. Increased alcohol sales were particularly noticeable for on-premise sales, strengthening the argument that the lowered drinking age was at least partially responsible for the observed changes.

Aggregate distilled spirits sales in 25 States were examined by Barsby and Marshall (1977). No significant effects of lowered legal drinking ages on spirit sales were found. The authors tempered their conclusions, however, by noting four limitations of their study. First, any change in distilled spirits consumption by youth following drinking age changes would have to be large before the impact would be seen in the aggregate figures (which include spirits consumed by drinkers of all ages). Second, little is known about changes in consumption patterns after legal changes; a change in location or quantity consumed per occasion as a result of the lowered drinking age, for example, could have significant adverse health consequences, independent of the total quantity consumed. Third, the analyses applied only to distilled spirits, not beer or wine, which are more popular beverages among young drinkers. Fourth, the time span covered by the study was

short, including only 1 year before and 1 year after the legal changes.

In their study, Douglass and Freedman (1977) avoided the last two design limitations by examining the monthly aggregate sales of draft beer, packaged beer, wine, and distilled spirits in Michigan over an 8-year period. A statistically significant increase in draft beer sales was associated with lowering the drinking age. The authors attributed the shift in draft beer sales to the lowered drinking age because no other confounding factors were identified that could have plausibly accounted for the observed relationship. No significant shifts were identified for any of the other beverage categories.

Per capita beer consumption in Canada was examined by Smart and Finley (1976). Eight provinces that lowered their drinking ages were compared with two that had not changed during the study period. Although increased beer sales were evident in the comparisons for three provinces that experienced a reduction in the drinking age, the increases were similar in magnitude to the experience of the two comparison provinces. Moreover, beer sales decreased in the other five provinces that had reduced the drinking age. The study's findings do not allow any general conclusion as to the effect of a lowered drinking age on total beer sales.

Smart (1977) compared sales of beer, wine, and distilled spirits in 25 States that reduced the drinking age with 25 States with unchanged drinking ages.[1] Although no significant differences between the two groups of States were identified for wine or distilled spirits, increases in beer sales were about 6 percent greater in the States with lowered drinking ages than in the States with an unchanged legal age.

Finally, Wagenaar (1982*a*,*b*) examined aggregate alcohol sales before and after the legal age was lowered and then later raised in Maine and Michigan. Results suggested significant effects of the legal age on aggregate alcohol sales, but interpretation was complicated by coincident implementation of other policies influencing alcohol sales, such as beverage container deposit laws.

In sum, a number of studies from the 1970's have used aggregate sales data to determine whether sales have increased following reductions in the minimum drinking age. Typically, results have shown that beer sales have increased following the enactment of a lower drinking age. Sales of wine and distilled spirits, however, have generally remained unchanged following a reduction in the drinking age.

Research conducted during the 1980's and 1990's has tended to employ stronger methodological designs than earlier research. This research has also tended to concentrate on instances in which the drinking age was raised rather than lowered. During this period, five survey studies have been conducted on the effects of the legal drinking age on drinking prac-

[1] *Interpretation of Smart's results is complicated by an apparent misclassification of States. Four States were analyzed as comparison States that did not change the drinking age, when in fact they did reduce the legal age and should have been included with the other 25 States that changed their laws.*

tices of general populations of youth, and six survey studies have been performed on the law's effects on more narrowly defined populations of college students. As was perhaps expected because few youth select distilled spirits as their alcoholic beverage of choice, Hoadley et al. (1984) found no effects of a raised drinking age on aggregate sales of distilled spirits. Smith et al. (1984) surveyed Massachusetts and New York 16- to 17-year-olds and found that increasing the legal age from 18 to 20 had "minimal effects" on drinking practices of those under 18. Coate and Grossman (1988) examined data from the second U.S. National Health and Nutrition Examination Survey and concluded that a national increase in legal age from 19 to 21 would decrease the prevalence of frequent drinkers by 21 percent. Williams and Lillis (1986) surveyed a probability sample of 16- to 20-year-old New York residents before and after the legal age was raised from 18 to 19 and found that drinking levels decreased significantly for 17- and 18-year-olds.

O'Malley and Wagenaar (1991) provided the most recent information on the effects of the legal age on alcohol use, using repeated cross sections and cohort panels of probability samples of 17- to 25-year-olds surveyed in the Monitoring the Future study. Combined across all States that increased the minimum drinking age from 18 (to 19, 20, or 21), there was a 13.3 percent decrease in drinking in the past 30 days (expressed as a percentage of the total standard deviation). Among States that increased the legal age from 18 to 21, the decrease in drinking was 28.2 percent. Furthermore, the decreases appeared immediately after the change in law, with very slight additional decreases in the succeeding 2 years.

In sum, the results of studies based on general populations of youths have been mixed, but several studies have indicated that youth consumption decreased following an increase in the minimum legal drinking age. This general conclusion should be qualified by noting that little change was detected in spirits sales (although youth generally prefer beer) and that little change was detected among youth who would have been too young to drink legally both before and after a change in the law.

In contrast to the general population youth samples in the five preceding studies, six studies used more limited samples of college students. These college student studies typically are based on surveys administered to undergraduates in introductory psychology or social science courses, and results of most of these studies indicate little effect of the legal drinking age on the reported drinking practices of these students. Hughes and Dodder (1986) surveyed college students in Oklahoma and found that the prevalence of drinkers decreased only one percentage point after the legal age was raised (from 88 percent before to 87 percent after). Engs and Hanson (1986, 1988) surveyed students at dozens of colleges and universities from 1982 to 1988. They found slightly lower rates of self-reported drinking problems in States with higher legal drinking ages, but conclude overall that collegiate drinking practices changed little

as a result of the increase in legal age. Mooney et al. (1988) compared samples of university students in Louisiana when the drinking age there was 18, with North Carolina, where the legal age was 21. They concluded that restricting access to alcohol does not lead to lower alcohol consumption, but drives young drinkers "underground." George et al. (1989) studied college students in New York before and after the age was increased from 19 to 21. While the law did affect the location of drinking (shifting drinking from public licensed establishments to private unlicensed locations), little effect on overall amounts of alcohol consumed was found. Gonzalez (1989) surveyed problem drinking among students at nine Florida universities. He concluded there was no appreciable effect of raising the legal age on quantity-frequency measures of drinking, but that there was a trend toward increased alcohol-related problems after the increase in the drinking age. Finally, Perkins and Berkowitz (1989) surveyed freshmen and sophomores at a single New York college before and after an increase in legal age from 18 to 19, concluding that there was little change in drinking or drinking problems. Interestingly, after the age was raised, 59 percent reported support for the higher age, despite the fact that 76 percent reported weekly drinking.

While many of these college student studies confirm the effects of the legal drinking age found in national samples, some do not. There are two possible reasons for divergent findings. First is the poor quality of the sampling methods employed in some studies. Students enrolled in introductory social science classes at selected colleges and universities are not representative of all youth of that age, or of all college students that age. Much more weight must be given to studies employing nationally representative probability samples of college-age youth. Second, it is clearly possible that the drinking age has different effects among specific pockets in the overall population. It may well be that among certain types of students at certain colleges, the increase in legal age had little effect. However, the overall effect for the Nation as a whole appears to be a decline in alcohol use among those under 21.

EFFECTS OF MINIMUM DRINKING AGE ON TRAFFIC CRASHES

Most research on the legal drinking age focuses on its effects on the involvement of young drivers in traffic crashes. Dozens of studies of the effects of lowering or raising the legal age published during the 1970's and early 1980's were reviewed previously (Wagenaar 1983). Results of that review are included here for completeness.

Williams et al. (1974) examined the frequency of fatal traffic crashes among 15- to 17-year-old and 18- to 20-year-old drivers in Michigan, Wisconsin, and Ontario, where the legal drinking age had been lowered. The frequency of fatal crashes in those areas for 3 years prior to and 1 year after the legal changes was compared with the frequency of fatal crashes in the contiguous States of Indiana, Illinois, and Minnesota. Drinking age laws in the comparison

States were not changed during the time period studied.

Significant increases in the number of fatal crashes were found among both 15- to 17-year-old and 18- to 20-year-old drivers in the jurisdictions that had lowered the legal drinking age. Separate analyses of single-vehicle and nighttime fatal crashes, a large proportion of which are known to be alcohol related, revealed larger increases than were found by analyses of all fatal crashes. Increased involvement in fatal crashes by youth was substantially greater for Michigan and Ontario than for Wisconsin. The smaller effect in Wisconsin was most likely a result of the less drastic change in the legal availability of alcohol. In Wisconsin, prior to the legal change, 18- to 20-year-olds could legally purchase beer; the new law simply extended that privilege to other types of alcoholic beverages.

Naor and Nashold (1975) also studied the impact of the legal change in Wisconsin on highway fatalities. Although the frequency of alcohol-related fatalities increased concomitantly with the legal change, the proportion of all fatally injured drivers with elevated blood-alcohol levels did not change significantly. Naor and Nashold used the latter finding to argue that the reduced drinking age had no effect on traffic crashes among youth. However, since beer, the beverage of choice among young people, was legally available prior to the drinking age change evaluated, this investigation alone should not be considered an adequate test of the effects of a lowered legal drinking age.

Cucchiaro et al. (1974) used monthly time series of traffic crashes to evaluate the impact of reducing the drinking age in Massachusetts. They examined crash involvement of drivers ages 15 to 17, 18 to 20, 21 to 23, and 24 and over. After the drinking age was lowered, 18- to 20-year-old drivers experienced significant increases in total fatal crashes, alcohol-related fatal crashes, and alcohol-related property damage crashes. None of the crash measures changed significantly for drivers aged 21 and over.

Douglass et al. (1974) also used monthly time series of traffic crashes to assess the effects of reduced drinking ages in Maine, Michigan, and Vermont. Collision involvement of young drivers in these States was compared with collision involvement of older drivers within the same State and with young drivers in Louisiana, Pennsylvania, and Texas, which held the drinking age constant over the study period. Analyses revealed significant increases in involvement in alcohol-related crashes among young drivers in both Michigan and Maine. No significant increases in alcohol-related crashes among youth were observed in any of the comparison States, nor were there any significant shifts for older drivers within Maine or Michigan. The researchers suggested that the lack of significant changes in the frequency of traffic crashes in Vermont, which also lowered its drinking age, may have been a result of the relative ease with which 18- to 20-year-olds in Vermont could obtain alcoholic beverages prior to the reduction of the drinking age by driving to New York, which had a drinking age of 18 at the time.

Douglass and Freedman (1977) replicated some of the earlier analyses, using 4

years of crash data after the legal age was lowered in Michigan. Results demonstrated that increased alcohol-related crashes among Michigan youth, identified in the 1974 research, persisted over the first 4 years with the lower drinking age (1972–75). Evaluation of the Michigan experience continued with analyses of fatal crashes in Michigan by Flora et al. (1978). Although these authors did not use the same analytical techniques used by Douglass, the impact of the 1972 reduction in the legal drinking age on alcohol-related traffic crashes among youth was again observed.

An increase in alcohol-related collisions following Ontario's reduction in drinking age was also reported by Schmidt and Kornaczewski (1975), who examined crash data from 1967 to 1971. The lack of monthly data and the failure to analyze separately only 18- to 20-year-old drivers made this study a conservative test of the effects of a reduced drinking age. Nonetheless, the researchers found a significant increase in crash involvement among drivers aged 16 to 19 after the law was changed.

Whitehead et al. (1975) examined crash involvement among drivers ages 16 to 20 and 24 in London, Ontario, for the period 1968–73. After Ontario's drinking age was lowered, alcohol-related crashes among drivers ages 18 to 20 increased 150 to 300 percent. In contrast, for the first year after the legal change, 24-year-old drivers experienced only a 20 percent increase in alcohol-related crashes, and their collision frequency returned to the prechange level the second year after the reduced drinking age took effect. In a followup study, Whitehead (1977) examined an additional 2 years of collision data. A total of 4 years of crash-involvement data after the reduction in drinking age demonstrated the permanence of higher rates of alcohol-related collisions among young drivers after a reduction in the minimum drinking age.

Warren et al. (1977) evaluated the impact of reduced drinking ages on traffic fatalities in Alberta, Manitoba, New Brunswick, and Saskatchewan. Only those fatalities for which a blood-alcohol concentration (BAC) test was administered were included in the analyses. Frequencies of alcohol-related fatalities for 15- to 20-year-old drivers before and after a reduction in drinking age were compared within each province. Some increases in fatalities among 15- to 20-year-old drivers were observed in the study jurisdictions at the time the drinking age was lowered. However, since the BAC legally defined as drunk driving in Canada was reduced at about the same time that the drinking age was lowered, the authors pointed out that effects of the BAC legislation were confounded with effects of the lower legal drinking age. Furthermore, insufficient numbers of prechange observations were available to control adequately for the random variation in the number of traffic fatalities from year to year. As a result, although increases in fatalities among youth occurred after the drinking age was lowered, the increases cannot be unequivocally attributed to the lower drinking age.

One of the provinces investigated by Warren, Saskatchewan, was also studied

by Shattuck and Whitehead (1976). After the drinking age was lowered from 21 to 19 in April 1970, 16- to 20-year-old drivers exhibited increases of 20 to 50 percent in alcohol-related crashes. After the drinking age was lowered from 19 to 18 in June 1972, this same group of drivers experienced further increases in alcohol-related crashes. Thus, two reductions in the legal age were associated with increased alcohol-related crashes among both the newly enfranchised drinkers and the underage population.

Bako et al. (1976) examined the frequency of drivers with BACs of 0.08 percent or greater among those fatally injured in the province of Alberta. An increase of 118 percent was observed in the incidence of fatal collisions related to alcohol among drivers aged 15 to 19 after the drinking age was lowered. The researchers concluded that their findings support the argument that lowered drinking ages lead to increased alcohol-related collisions among youth.

The reduction in the legal drinking age for beer and wine in Illinois (from 21 to 19) was evaluated by the Illinois Department of Transportation (1977). Comparisons between numbers of crash fatalities in Illinois and five companion States were used as the basis for the conclusion that the lowered drinking age in Illinois resulted in a small increase in fatalities among drivers ages 19 and 20.

The Alcohol Epidemiologic Data System (AEDS; 1980) examined annual traffic fatality counts in the 50 States from 1970 to 1978. Although increased motor vehicle deaths were seen in some States following a reduction in drinking age, it was concluded that "differences in highway fatalities with the change in drinking laws do not appear significantly large" (AEDS 1980, p. 1). The authors readily admit, however, that their analyses did not adequately control for the effects of numerous confounding factors that occurred in the 1970's, such as fuel shortages, speed limit reductions, and so on.

After Alabama lowered its drinking age from 21 to 19 in 1975, alcohol-related crashes increased significantly among drivers aged 18 to 20 according to Brown and Maghsoodloo (1981). Koch (1981) pointed out the methodological limitations of Brown and Maghsoodloo's study and argued that it did not establish a causal connection between the drinking age and crashes. The study did not include adequate comparisons between alcohol-related and nonalcohol-related crashes and between States that changed the drinking age and those that did not. However, the association between lowered drinking age and increased alcohol-related crashes in Alabama is consistent with the results from studies of other States and provides additional support for the conclusion that lower drinking ages increase alcohol-related crashes.

Lynn (1981) analyzed annual counts of alcohol-related crashes in Virginia from 1969 to 1979. Drivers ages 16 to 20 experienced an identifiable increase in involvement in alcohol-related crashes beginning in 1974, when the drinking age for beer was lowered to 18. In contrast, drivers age 25 and older experienced a decrease in alcohol-related crashes during the same

period. Lynn concluded that the lowered drinking age was responsible for increased alcohol-related crashes among young drivers in Virginia in the late 1970's and recommended a gradual return to the legal drinking age of 21.

Cook and Tauchen (1982) estimated the effect of providing youth legal access to alcohol through lowering the drinking age, using nationwide U.S. data on the number of young people killed in motor vehicle crashes. Analyses of fatality data pooled across States and years revealed that a 7 percent increase in the number of youth killed in automobile crashes was associated with reductions in the legal age. The 7 percent estimate applies to youth killed in all motor vehicle crashes, whether drivers or passengers, drinking or not drinking, at the time of the crash. Crashes in which a young driver contributed to the death of an older person, but in which a young person was not killed, were not taken into account by Cook and Tauchen. Some evidence of a spillover effect of lower drinking ages was also found. An estimated 6 percent increase in deaths among 16- to 17-year-olds was associated with reductions in the drinking age.

The literature reviewed here shows that most of the investigations of the impact of lower legal drinking ages on involvement in motor vehicle collisions have found significant increases in crash involvement among previously underage drivers who acquired the right to drink under the new laws (usually 18- to 20-year-old drivers). Several studies have also demonstrated substantial increases in crash involvement among underage drivers (usually 16 and 17 years old) following reductions in the minimum drinking age. Many of the evaluations of lowered drinking ages have methodological limitations, such as (1) lack of adequate comparison groups, (2) lack of extended longitudinal data, and (3) outcome measures of limited coverage with small numbers of cases (for example, only fatal crashes or only crashes for which a chemical BAC test was administered). The pattern of results across studies of numerous jurisdictions, however, using a variety of analytic methods, leads to the conclusion that lowered drinking ages typically resulted in an increase in problems related to highway safety among young drivers.

The view that lower legal drinking ages increase the involvement of youth in crashes was not universally held in the 1970's. Zylman was a well-known opponent of a causal interpretation of the observed relationships. Zylman (1973, 1974, 1976, 1977) criticized several of the studies reviewed above. He argued that observed increases in involvement in alcohol-related crashes among youth after the drinking age was lowered were not the result of the changes in drinking age, but rather were a result of (1) random fluctuations in traffic crash time series, (2) continuation of trends of increasing alcohol consumption and alcohol-related crashes among youth that were evident prior to the legal changes, or (3) increased attention to alcohol-related traffic offenses by law enforcement officers. A careful review of the research, however, indicates that these factors were taken into account in

the more sophisticated investigations and that the effect of the drinking age was still evident. For example, studies that explicitly controlled for both long-term trends and random fluctuations found significant effects of the lower drinking age. Although Zylman correctly points out limitations of relying solely on analyses of police-reported alcohol involvement, the effects of lowered drinking ages (although of smaller magnitude than analyses based on police reports) have been observed using alternative measures of alcohol involvement not influenced by the reporting practices of police (such as analyses of single-vehicle, nighttime, and weekend crashes).

In addition to the studies above, which have evaluated the consequences of lowering the drinking age, numerous other studies have evaluated the effects of raising the drinking age. Roy and Greenblatt (1979) compared the number of teenagers charged with driving under the influence of liquor (DUIL) who appeared in Massachusetts courts before the legal age was raised with similar data for a 1-month period after the drinking age change.[2] Small increases in youth DUIL arrests were used to conclude that the raised drinking age led to increased drinking-and-driving problems among youth. This study, however, is flawed in the following respects: (1) DUIL arrests alone are an inadequate measure of alcohol-related highway safety hazards, because young drivers are more likely than older drivers to be involved in an alcohol-related crash but are less likely to be arrested for DUIL (Organization for Economic Cooperation and Development, 1975);[3] (2) the design is a single-group pre-post test, inherently characterized by low internal validity because of its lack of comparison groups and an extended series of observations (Cook and Campbell 1979); (3) related to the basic inadequacy of the research design is the lack of any statistical control on time-ordered trends, seasonal cycles, or random fluctuations in the frequency of DUIL arrests.

Hingson et al. (1983) also evaluated the effects of the increase from 18 to 20 in the drinking age in Massachusetts (effective April 1979). Traffic fatality data from April 1976 through April 1981 were analyzed for Massachusetts and compared with similar data for New York, which had not raised the drinking age. After the drinking age was raised, fatal crashes among 18- to 19-year-old Massachusetts drivers decreased 15 percent, versus a 16-percent increase among 18- to 19-year-olds in New York. A similar effect was not found for 16- to 17-year-old drivers. Telephone surveys of Massachusetts and New York youth before and after the new law was implemented revealed decreases in self-reported nonfatal crash involvement in both States. Analyses of 16- to 19-year-olds as a single group indicated

[2] *The legal drinking age in Massachusetts was raised from 18 to 20 on April 16, 1979. The pre-post comparison was the number of DUIL arrests in February 1979 versus the number in October 1979.*

[3] *In addition, subjective considerations are more likely to influence whether a particular drinking driver is arrested than whether a drinking driver is involved in a crash.*

that the decrease in Massachusetts was not significantly greater than the decrease in New York. Responses to questions concerning the frequency of driving after drinking revealed no change in driving after heavy drinking that was attributable to the higher drinking age. However, the proportion of 16- to 19-year-old Massachusetts youth who reported driving after any drinking (as opposed to driving after heavy drinking) decreased from 51 percent before the raised drinking age to 42 percent after; no decrease was found in New York. Separate results for the focal group of interest, 18- to 19-year-olds, were not reported.

Voas and Moulden (1980) briefly reported analyses of the effect of raising the drinking age in Maine from 18 to 20. Their initial analyses of 18-year-old drivers indicated larger reductions in single-vehicle nighttime crashes by men (many of which involve alcohol) than in single-vehicle daytime crashes (few of which involve alcohol). They estimated an average reduction of 2.4 crashes per month among 18-year-old drivers over the first year with the new law. No effect was found for 19-year-olds. The authors pointed out that these results were based on early analyses for a short postchange period in only one small State and that additional results from other States are needed before any general conclusions on the effects of higher drinking ages are made.

Klein (1981) also evaluated the effect of Maine's increase in the drinking age. Involvement in nighttime crashes among males and in single-vehicle nighttime crashes among males were used as proxies for alcohol-related crashes. Time-series analyses using data from 1974 to 1979 revealed significant reductions in crash involvement from 14 to 20 percent after Maine's higher legal age was implemented.

Maxwell (1981) conducted similar time-series analyses on Illinois data from 1977 to 1980. A significant 9 percent decrease in involvement in single-vehicle nighttime crashes by males was found among 19- to 20-year-old drivers after Illinois raised the drinking age from 19 to 21 (effective January 1980). In contrast, no significant changes in these measures were found for drivers under the age of 19 or over the age of 20.

Filkins and Flora (1981) examined random samples of crash-involved youth in Michigan. After the drinking age was raised from 18 to 21, significant reductions of 11 to 24 percent were found in the rate of involvement in alcohol-related crashes among 18- to 21-year-old drivers. Separate analyses of fatal crashes, however, identified no significant changes associated with the higher drinking age. The authors concluded that the overall results clearly demonstrated that the legal drinking age influences the amount of alcohol-impaired driving among young people.

Wagenaar (1981) analyzed a 20 percent random sample of all reported motor vehicle crashes in Michigan between January 1972 and December 1979. Using a multiple time-series design, the frequency of alcohol-related crashes among young drivers was compared with the frequency of nonalcohol-related crashes, and crash involvement of young drivers was compared with that of older drivers.

Results showed that an estimated 18-percent reduction in involvement in alcohol-related crashes among young drivers was associated with Michigan's increase in the drinking age.

Williams et al. (1981) analyzed fatal crash involvement from 1975 to 1980 in nine States that raised the drinking age, comparing them with neighboring States with unchanged drinking ages during the period studied. Two of the States raised the drinking age from 18 to 19 and included grandfather clauses that permitted those already aged 18 to continue to drink; crash data in these States were excluded for the first 12 months after the legal changes. Eight of the nine States experienced decreases in youth involvement in nighttime fatal crashes after the drinking age was raised; crash reductions ranged from 6 to 75 percent. The smallest estimated reduction applied to Massachusetts, for which a nonsignificant 6-percent reduction in youth involvement in nighttime fatal crashes was found. Averaging across the nine States examined, the authors concluded that raising the drinking age in any given State should result in a 28-percent reduction in involvement in nighttime fatal crashes among the age group affected by the legal change.

Vingilis and Smart (1981) assessed the initial effects of Ontario's increase in the drinking age from 18 to 19. The law included a grandfather clause, permitting those age 18 at the time of the implementation of the higher legal age to continue to consume alcoholic beverages. Aggregate analyses of drinking-and-driving convictions and fatal crash involvement among 16- to 21-year-olds for a multiyear period revealed no significant changes associated with the increase in drinking age. These results should be interpreted cautiously for several reasons. First, only a small effect of the new law was expected, because the legal age was increased by only 1 year, with a grandfather clause that excluded 18-year-olds who already had the right to drink from the provisions of the new law. With a grandfather clause in the law, and analyses of only the first year under the new law, only a small proportion of 18-year-old Ontario drivers were denied easy access to alcohol during the study period. Finally, convictions and fatalities, being relatively rare events, are measures of limited sensitivity and respond only to major shifts in drinking-and-driving behavior of a sizable proportion of the population. Because of these factors it is important to consider these results preliminary findings, as Vingilis and Smart pointed out.

The NHTSA (1982) compared serious injury rates (including both motor vehicle and nonmotor vehicle injuries) across States with low drinking ages (18 or 19) and States with high drinking ages (20 or 21), using hospital emergency room data. States with higher drinking ages had lower nighttime serious injury rates among youth under 21 years old than did States with lower drinking ages.

Since the early 1980's, 29 additional studies have been conducted of the effects on crash involvement of policy changes that raised the legal age (table 1). Of the 29 studies, 20 conclude that policy

TABLE 1

U.S. studies of effects of raising the legal drinking age on traffic crashes, 1981-1992

Study	Jurisdiction	Data	Results	Conclusion
Vingilis and Smart 1981	Ontario	Drunk driving convictions and fatalities	n.s.	Equivocal
Maxwell 1981	Illinois	Single vehicle male crashes	–8.8%	Yes
Colon and Cutter 1983	50 States	Fatalities	n.s.	No
Dunham and Detmer 1983	Wisconsin	Crashes	–20%	Yes
Schroeder and Meyer 1983	Illinois	Fatalities	–7%	Yes
Hingson et al. 1983	Massachusetts	Single vehicle nighttime fatalities	n.s.	Equivocal
Colon 1984	50 States	Single vehicle fatalities		Equivocal
Smith et al. 1984	Massachusetts	Fatalities among 16- to 17-year-olds	n.s.	Equivocal
Arnold 1985	13 States	Fatalities	–13%	Yes
Choukroun et al. 1985	14 States	Fatalities		Equivocal
Asch and Levy 1987*a*	50 States	Fatalities	n.s.	No
DuMouchel et al. 1987	50 States	Fatalities	–9%	Yes
Hoskin et al. 1986	10 States	Single vehicle nighttime fatalities	–21%	Yes
Hughes and Leung 1986	Wisconsin	Crashes		Yes
MacKinnon and Woodward 1986	3 States	Fatalities	–11 to –18%	Yes
Wagenaar 1986	Michigan	Single vehicle nighttime crashes	–16%	Yes
Wagenaar and Maybee 1986	Texas	Injury crashes	–11%	Yes
Asch and Levy 1987*b*	47 States	Fatalities		Equivocal
Coate and Grossman 1987	50 States	Fatalities	– 8%	Yes
Hoxie and Skinner 1987	50 States	Fatalities	–11%	Yes
Loeb 1987	50 States	Fatalities	n.s.	No
Saffer and Grossman 1987*a,b*	50 States	Fatalities	– 8%	Yes
Wilkinson 1987	50 States	Fatalities	–1.2%	Yes
Decker et al. 1988	Tennessee	Fatalities	– 38%	Yes
Hinshaw and Mounce 1989	Texas	Crashes	–12%	Yes
Robertson 1989	30 States	Fatalities		Yes
Womble 1989	50 States	Fatalities	–12%	Yes
O'Malley and Wagenaar 1991	42 States	Single vehicle nighttime fatalities	–15%	Yes
Jones et al. 1992	50 States	Fatalities	–4%	Yes

changes increasing the legal age for drinking reduce youth crash involvement. Several provide equivocal conclusions, and only three conclude that the raised legal age did not affect youth crash involvement. Results from O'Malley and Wagenaar's recent (1991) nationwide study are presented here in some detail.

We conducted analyses of official statistics on fatal automobile crashes (using the Fatal Accident Reporting System) that were coordinated with analyses of the Monitoring

the Future nationwide survey data. Both self-report and crash data were examined for a subset of States with crash data available for 3 years before and after a legal age change. Thirteen States met this requirement: Delaware, Florida, Georgia, Illinois, Maryland, Massachusetts, Michigan, Nebraska, New Jersey, Ohio, Oklahoma, Tennessee, and Texas. Our objective was to compare findings from analyses of crashes with self-report data for the same States during the same time period, to determine whether declines in crashes can be attributed to lower rates of alcohol use.

We estimated effects of the raised legal age on rates of fatal crashes for the following four categories: (1) single-vehicle nighttime (SVN) crashes involving a driver less than 21 years old, (2) SVN crashes involving a driver 21 years old or older, (3) daytime crashes involving a driver less than 21 years old, and (4) daytime crashes involving a driver 21 or older.

The principal finding was that there was a significant decline in SVN fatal automobile crashes among drivers less than 21 years of age following increases in minimum drinking age. The largest rate change occurred among States whose change in minimum drinking age was 3 years—that is, from 18 to 21. In these States, there was a decline of 26.3 percent in the rate per licensed driver of SVN fatal crashes involving drivers under 21 (and a decline of 27.8 percent in the frequency of such crashes), comparing the 3 years before and the 3 years after a law change, as estimated by Box-Jenkins time-series intervention models. States with a 2-year change (from 18 to 20 or from 19 to 21), and States with a 1-year change (from 18 to 19, from 19 to 20, or from 20 to 21) also showed significant declines in crash rates (18.6 and 21.6 percent, respectively).

To rule out other potential explanations for the decline in crashes following an increase in minimum drinking age, we conducted additional multivariate analyses utilizing the self-report data. These analyses controlled for sex, race, number of parents in home, parental education, college plans, religious commitment, truancy, number of evenings out per week, and weekly income, in addition to the number of miles driven per week. A significant effect of minimum drinking age remained after controlling all these other variables, and the magnitude of the drinking age effect was only slightly reduced after controlling all these variables.

In sum, studies of traffic crash data contain substantial support for the relationship between the legal drinking age and the involvement of young drivers in traffic crashes. With occasional exceptions, studies of reductions in the legal drinking age have indicated that these reductions in the legal drinking age were accompanied by increases in crash and fatality data among youth in the affected ages. Similarly, the preponderance of studies of increases in the legal drinking age have indicated that such increases in the legal drinking age have had the effect of reducing traffic crashes and fatality rates among youth in the affected ages.

EFFECTS OF MINIMUM DRINKING AGE ON OTHER HEALTH AND SOCIAL PROBLEMS

Only four studies have explicitly examined the effects of raising the legal drink-

ing age on other social and health problem indicators. The New York State division of alcoholism and alcohol abuse (1984) collected data on vandalism arrests in eight States and found that the four States that raised the legal age experienced a 16 percent decrease in vandalism rates. Hingson et al. (1985) examined nonmotor-vehicle trauma, suicide, and homicide fatalities before and after Massachusetts raised its legal age from 18 to 20; no significant reductions were found. However, Smith (1986) found increases of 22 to 44 percent in nontraffic trauma hospital admissions associated with a reduction in the legal drinking age in two Australian States. Most recently, Jones et al. (1992) examined mortality data for all 50 States and found that the drinking age of 21 was associated with significant reductions in deaths due to suicides, motor vehicle crashes, pedestrian injuries, and other unintentional injuries. Thus, three of four studies report significant effects of raising the drinking age in reducing social and health problems other than traffic crashes. One reason for Hingson and coworkers' divergent findings may be related to the large proportion of the Massachusetts population living within driving distance of bordering States with lower legal ages at the time of the study, as well as very little enforcement of the minimum age in Massachusetts.

RESEARCH NEEDS ON YOUTH ALCOHOL AVAILABILITY

Research has demonstrated long-term beneficial effects of raising the legal drinking age to 21. Much remains to be learned, however, about all the other many dimensions of the availability of alcohol to youth, including sources of alcohol, locations and patterns of youth drinking occasions, commercial and noncommercial providers of alcohol to youth, dimensions of the social environment that support youth drinking, and methods to modify public policy and mobilize communities to address issues of access to alcohol among youth.

Little is known about the nature of enforcement actions taken by State alcoholic beverage control agencies against retail outlets that sell alcohol to youth. How much effort is allocated to detection of underage sales? What methods are used to detect such sales? What consequences typically ensue for an outlet selling to youth? Do outlet owners and managers perceive a risk of detection for underage sales? How severe do they believe the consequences will be? Do current statutes and levels of enforcement serve a deterrence function? How do local police enforcement personnel perceive the problem of youth drinking? How do they respond to community pressures regarding enforcement priorities? What community actions are necessary to obtain enforcement attention to this issue?

Studies using several methodological approaches are necessary to answer continuing questions about youth alcohol availability. Time-series studies of available records using appropriate analytic techniques are needed for quasi-experimental evaluation of the effects of any future natural experiments with altered alcohol accessibility to youth, whether

these occur at national, State/provincial, or local levels. Several European countries are closely examining the evidence regarding the effects of changes in the minimum legal age in North America, and proposals to increase the legal age in those countries are receiving increasing attention. If such increases are implemented, they should be used as opportunities to replicate the studies in the United States, demonstrating the extent to which the U.S. experience is generalizable to other countries. Given significant differences across countries in terms of drinking cultures, alcohol distribution systems, and other characteristics, effects of the legal drinking age may differ.

While natural experiments can be put to good use in quasi-experimental studies, we should also move ahead to the next step—randomized community trials of efforts to reduce youth alcohol availability. With adequate resources, such randomized trials are feasible, particularly if substantial numbers of communities of modest size are included in the trial. The key factor in designing such studies for adequate power to detect plausible effect sizes is the number of communities (not the number of respondents within each community). Thus, large-scale studies of one to two dozen communities will be necessary (Wagenaar and Wolfson 1992).

New measurement methods are also needed, in addition to conventional measures based on available records and population surveys. We need observation studies of youth drinking, the locations in which such drinking occurs, and the situational and contextual factors that contribute to or impede youth alcohol use. Only a few observational studies of youth drinking have appeared. Most of these studies focus on the effects of demographics, group size, crowdedness, and music volume on drinking rates; none have examined the role of alcohol accessibility (Knibbe et al. 1991, 1992; van de Goor et al. 1990; Aitken 1985; Aitken and Jahoda 1983; Rosenbluth et al. 1978). Yet more than a third of high school seniors binge drink (five or more drinks on one occasion) regularly (Johnston et al. 1991). We know little about the physical and social environments where that drinking occurs. Where do youths binge drink? With whom? Where do they acquire the necessary alcoholic beverages? What are the social functions of such drinking occasions? Focus group studies are suggesting numerous dimensions of youth alcohol access and drinking practices that deserve detailed attention in larger controlled studies (Wagenaar et al. in press). Studies are needed that use multiple data collection strategies, including focus group and other qualitative methods to initially identify potentially important dimensions; surveys of probability samples of youth regarding the prevalence, frequency, and intensity of various practices; and structured observation of public places in which youth drink. Finally, cohort studies of youth starting at ages 10 to 12, when most youth first drink alcohol, and continuing through age 25 or 30, after which drinking rates and drinking problems significantly diminish, would contribute much to our understanding of the development of particular drinking practices and problems.

In addition to survey and observational studies of youth, direct purchase attempt studies of commercial alcohol outlets are needed. Only three such studies have been reported to date. McKnight (1990) found sales refused to underage youth in only one-third of 100 establishments tested at 8 sites around the United States. The Insurance Institute for Highway Safety (Preusser and Williams in press) found that 97 of a sample of 100 purchase attempts in Washington, DC, and 124 of 200 attempts in New York resulted in the sale of alcohol to 17- and 18-year-olds. Finally, in northern Minnesota successful alcohol purchases were made in 47 percent of 336 attempts at liquor stores (Forster et al. 1991). We have a much larger study including 1,800 purchase attempts in 21 cities in Minnesota and Wisconsin currently in progress (Wagenaar et al. 1991). Clearly, results to date across several studies suggest that sales to minors are widespread, not limited to a small number of atypical outlets. Despite these important studies, many questions remain. These initial studies show high rates of illegal sale of alcohol to youth. Replication is needed with diverse sets of outlets across geographic areas. Do outlets that sell only beer and wine have a higher or lower propensity to sell to underage youth? What about outlets that sell only beer that has a lower alcohol content (e.g., 3.2 percent)? Is on-premise service of alcohol to youth a major source of sales? Do convenience stores sell at higher or lower rates to underage youth? Clearly, additional analyses are needed on what types of outlets and what types of clerks are most likely to sell alcohol to youth. Such information will directly inform intervention efforts to reduce the accessibility of alcohol to youth. For example, if certain types of outlets predominate in sales of alcohol to minors, enforcement actions and other prevention efforts could be targeted to those kinds of outlets.

In conclusion, the literature on the effects of minimum legal drinking age policies is extensive and demonstrates significant reductions in drinking rates and alcohol-related problems as a result of raising the legal age from 18 to 21 in the United States. However, as noted at the beginning of this paper, most youth continue to have access to alcohol, most drink at least occasionally, and a substantial fraction regularly become intoxicated. Social costs from injuries, deaths, and damage associated with underage drinking remain high. Continuing research and programmatic efforts are needed to fully implement the "age 21" policy, as well as to change the broader social environment that makes alcohol appealing to youth.

ACKNOWLEDGMENTS

Analyses of the Monitoring the Future data were conducted collaboratively with Patrick O'Malley of The University of Michigan Institute for Social Research; for full details, readers are referred to O'Malley and Wagenaar (1991). Warm thanks to Jan Howard and Mike Hilton for helpful comments on an earlier draft and to Gary Hornseth for assistance with reviewing the literature.

REFERENCES

Aitken, P.P. An observational study of young adult's drinking groups. II. Drinking purchasing procedures, group pressures and alcohol consumption by companions as predictors of alcohol consumption. *Alcohol Alcohol* 20(4):445–457, 1985.

Aitken P.P., and Jahoda, G. An observational study of young adult's drinking groups. II. Drink preferences, demographic and structural variables as predictors of alcohol consumption. *Alcohol Alcohol* 18(2):135–150, 1983.

Alcohol Epidemiologic Data System. Data Computations on the Influence of Changes in State Drinking Age Laws on Highway Fatalities in the Deaths of the 70s. Working paper 37. Gaithersburg, MD: General Electric Co., 1980.

Arnold, R.D. *Effect of Raising the Legal Drinking Age on Driver Involvement in Fatal Crashes: The Experience of Thirteen States.* NHTSA Technical Report (DOT-HS-806-902). Springfield, VA: National Technical Information Service, 1985.

Asch, P., and Levy, D.T. Does the minimum drinking age affect traffic fatalities? *J Policy Anal Manage* 6(2):180–192, 1987*a*.

Asch, P., and Levy, D.T. Young Driver Fatalities: The Roles of Drinking Age and Drinking Experience. Working paper. Rutgers University, 1987*b*.

Baker, S.P.; O'Neill, B.; and Karph, R.S. *The Injury Fact Book.* Lexington, MA: D.C. Heath and Co., 1984. 344 pp.

Bako, G.; MacKenzie, W.C.; and Smith, E.S.O. The effect of legislated lowering of the drinking age on total highway accidents among young drivers in Alberta, 1970-1972. *Can J Public Health* 67(2):161–163, 1976.

Barsby, S.L., and Marshall, G.L. Short-term consumption effects of a lower minimum alcohol purchasing age. *J Stud Alcohol* 38(9):1665–1679, 1977.

Brown, D.B., and Maghsoodloo, S. A study of alcohol involvement in young driver accidents with the lowering of the legal age of drinking in Alabama. *Accident Anal Prev* 13(4):319–322, 1981.

Choukroun, J.M.; Ravn, I.; and Waner, C. The Relationship Between Increases in Minimum Purchase Age for Alcoholic Beverages and the Number of Traffic Fatalities. University of Pennsylvania Social Systems Sciences Department, 1985.

Coate, D., and Grossman, M. Change in alcoholic beverage prices and legal drinking ages: Effects on youth alcohol use and motor vehicle mortality. *Alcohol Health Res World* 12(1):22–25, 1987.

Coate, D., and Grossman, M. Effects of alcoholic beverage prices and legal drinking ages on youth alcohol use. *J Law Econ* 31(1):145–171, 1988.

Colon, I. The alcohol beverage purchase age and single-vehicle highway fatalities. *J Safety Res* 15(4):159–162, 1984.

Colon, I., and Cutter, H.S.G. The relationship of beer consumption and state alcohol and motor vehicle policies to fatal accidents. *J Safety Res* 14(2):83–89, 1983.

Cook, T.D., and Campbell, D.T. *Quasi-experimentation: Design and Analysis Issues for Field Settings.* Chicago: Rand McNally, 1979. 405 pp.

Cook, P.J., and Tauchen, G. The Effects of Minimum Drinking Age Legislation on Youthful Auto Fatalities, 1970–1977. Unpublished manuscript, 1982.

Cucchiaro, S.; Ferreira, Jr., J.; and Sicherman, A. *The Effect of the 18-year-old Drinking Age on Auto Accidents.* Cambridge, MA: Massachusetts Institute of Technology, Operations Research Center, 1974. 32 pp.

Decker, M.D; Graitcer, P.L.; and Schaffner, W. Reduction in motor vehicle fatalities associated with an increase in the minimum drinking age. *JAMA* 260(24):3604–3610, 1988.

Douglass, R.L.; Filkins, L.D; and Clark, F.A. *The Effect of Lower Legal Drinking Ages on Youth*

Crash Involvement. The University of Michigan Highway Safety Research Institute, DOT Report No. DOT HS-801213, PB No. 325917, 1974.

Douglass, R.L., and Freedman, J.A. *Alcohol-related Casualties and Alcohol Beverage Market Response to Beverage Alcohol Availability Policies in Michigan.* 2-volume technical report #UMH-SRI7737. The University of Michigan Highway Safety Research Institute, August 1977.

Driver, H.E., and Swann, P.F. Alcohol and human cancer (Review). *Anticancer Res* 7(3A):309–320, 1987.

Dryfoos, J.G. *Adolescents at Risk: Prevalence and Prevention.* New York: Oxford University Press, Inc., 1990. 280 pp.

DuMouchel, W.; Williams, A.F.; and Zador, P.L. Raising the alcohol purchase age. *J Legal Stud* 16(1):249–266, 1987.

Dunham, N.C., and Detmer, D.E. Traffic accidents and the legal drinking age in Wisconsin. *Wis Med J* 82(3):11–13, 1983.

Engs, R.C., and Hanson, D.J. Age-specific alcohol prohibition and college students' drinking problems. *Psychol Rep* 59(2):979–984, 1986.

Engs, R.C., and Hanson, D.J. University students' drinking patterns and problems: Examining the effects of raising the purchase age. *Public Health Rep* 103(6):667–673, 1988.

Filkins, L.D., and Flora, J.D. *Alcohol-related Accidents and DUIL Arrests in Michigan: 1978-1979.* Interim Report. The University of Michigan Highway Safety Research Institute, Report #UMHSRI8124, 1981.

Flora, J.D.; Filkins, L.D.; and Compton, C.P. *Alcohol Involvement in Michigan Fatal Accidents: 1968–1976.* Interim Report. The University of Michigan Highway Safety Research Institute, Report #UMHSRI7816, 1978.

Forster, J.L.; Wagenaar, A.C.; Perry, C.L.; and Anstine, P.S. "Alcohol Availability to Underage Youth: Rates of Off-sale Liquor Licensee Sales to Minors." Paper presented at the 119th annual meeting of the American Public Health Association, Atlanta, Georgia, November 10-14, 1991.

George, W.H.; Crowe, L.C.; Abwender, D.; and Skinner, J.B. Effects of raising the drinking age to 21 years in New York State on self-reported consumption by college students. *J Appl Soc Psychol* 19(8):623–635, 1989.

Gonzalez, G.M. Early onset of drinking as a predictor of alcohol consumption and alcohol-related problems in college. *J Drug Educ* 19(3):225–230, 1989.

Grant, B.F.; Dufour, M.C.; and Harford, T.C. Epidemiology of alcoholic liver disease. *Semin Liver Dis* 8(1):12–25, 1988.

Hingson, R.; Merrigan, D.; and Heeren, T. Effects of Massachusetts raising its legal drinking age from 18 to 20 on deaths from teenage homicide, suicide, and nontraffic accidents. *Pediatr Clin North Am* 32(1):221–233, 1985.

Hingson, R.W.; Scotch, N.; Mangione, T.; Meyers, A.; Glantz, L.; Heeren, T.; Lin, N.; Mucatel, M.; and Pierce, G. Impact of legislation raising the legal drinking age in Massachusetts from 18 to 20. *Am J Public Health* 73(2):163–170, 1983.

Hinshaw, W.M., and Mounce, N.H. An Evaluation of the Effectiveness of Raising the Minimum Drinking Age in Texas. Texas Transportation Institute, Texas A&M University System, September 1989.

Hoadley, J.F.; Fuchs, B.C.; and Holder, H.D. The effect of alcohol beverage restrictions on consumption: A 25-year longitudinal analysis. *Am J Drug Alcohol Abuse* 10(3):375–401, 1984.

Hoskin, A.F.; Mathews, D.Y.; and Carraro, B.A. The effect of raising the legal minimum drinking age on fatal crashes in 10 States. *J Safety Res* 17(3):117–121, 1986.

Hoxie, P., and Skinner, D. *A Statistical Analysis of the Effects of a Uniform Minimum Drinking*

Age. Technical Report (DOT-HS-807-082, DOT-TSC-NHTSA-87-1). Springfield, VA: National Technical Information Service, 1987.

Hughes, S.P., and Dodder, R.A. Raising the legal minimum drinking age: Short-term effects with college student samples. *J Drug Issues* 16(4):609–620, 1986.

Hughes, D.J., and Leung, K.S. "Measuring the Impact of Wisconsin's Age 19 Drinking Law." Paper presented at the Transportation Research Board Annual Meeting, January 1986.

Illinois Department of Transportation, Division of Traffic Safety. *Assessment of the Effect on Traffic Safety of Lowering the Legal Drinking Age in Illinois.* November 1977.

Johnston, L.D.; O'Malley, P.M.; and Bachman, J.G. *Drug Use Among American High School Seniors, College Students and Young Adults, 1975-1990.* DHHS Pub. No. (ADM)91-1813. National Institute on Drug Abuse, Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1991.

Jones, N.E.; Pieper, C.F.; and Robertson, L.S. The effect of legal drinking age on fatal injuries of adolescents and young adults. *Am J Public Health* 82(1):112–115, 1992.

Kandel, D.B. Issues of sequencing of adolescent drug use and other problems. *Drugs Society* 3(1-2):55–76, 1989.

Kandel, D.B. *The Stage Theory of Drug Consumption.* Ann Arbor: University of Michigan Press, 1989.

Klein, T. *The Effect of Raising the Minimum Legal Drinking Age on Traffic Accidents in the State of Maine.* U.S. National Highway Traffic Safety Administration, Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1981.

Knibbe, R.A.; Oostveen, T.; and van de Goor, I. Young people's alcohol consumption in public drinking places: Reasoned behavior or related to the situation? *Br J Addict* 86(11):1425–1433, 1991.

Knibbe, R.A.; van de Goor, I.; and Drop, M.J. "Contextual Influences on Young People's Drinking Rates in Public Drinking Places: An Observational Study." Paper presented at the 18th Annual Alcohol Epidemiology Symposium, Toronto, Ontario, Canada, June 1–5, 1992.

Koch, G.G. Commentary on statistical issues underlying the evaluation of a social policy change on motor vehicle accident involvement. *Accident Anal Prev* 13(4):323–329, 1981.

Kusserow, R.P. *Youth and Alcohol: A National Survey.* Washington, DC: Office of the Inspector General, U.S. Department of Health and Human Services, 1991.

Loeb, P.D. The determinants of automobile fatalities. *J Transport Econ Policy* 21(3):279–287, 1987.

Lynn, C. *The Effects of Lowering the Legal Drinking Age in Virginia.* Charlottesville, VA: Virginia Highway and Transportation Research Council, 1981.

MacKinnon, D.P., and Woodward, J.A. The impact of raising the minimum drinking age on driver fatalities. *Int J Addict* 21(12):1331–1338, 1986.

Maisto, S.A., and Rachal, J.V. Indications of the relationship among adolescent drinking practices, related behaviors, and drinking age laws. In: Wechsler, H., ed. *Minimum-Drinking-Age Laws.* Lexington, MA: D.C. Heath and Co., 1980. pp. 155–176.

Maxwell, D.M. *Impact Analysis of the Raised Legal Drinking Age in Illinois.* U.S. National Highway Traffic Safety Administration, Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1981.

McFadden, M., and Wechsler, H. Minimum drinking age laws and teenage drinking. *Psychiatr Opinion* 16(3):22–23, 26–28, 1979.

McKnight, J.A. Intervention with alcohol-impaired drivers by peers, parents and purveyors of alcohol. *Health Educ Res* 5(2):225–236, 1990.

Mooney, L.A.; Gramling, R.; and Forsyth, C. "Legal Drinking Age and Alcohol Consumption:

Deterrence or Amplification." Paper presented at the annual meeting of the Society for the Study of Social Problems, Atlanta, Georgia, 1988.

Naor, E.M., and Nashold, R.D. Teenage driver fatalities following reduction in the legal drinking age. *J Safety Res* 7(2):74–79, 1975.

National Highway Traffic Safety Administration. *Evaluation of Minimum Drinking Age Laws Using the National Electronic Injury Surveillance System.* Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1982. (Copies can be obtained from the National Technical Information Service, 5285 Port Royal Road, Springfield, VA 22161, (703)487-4650, Pub. No. PB 82189010.)

National Highway Traffic Safety Administration. *Alcohol and Highway Safety 1984.* Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1985.

New York State Division of Alcoholism and Alcohol Abuse. *Drinking Age Laws and Vandalism Arrests: Preliminary Findings 1976–1981.* May 1984.

O'Malley, P., and Wagenaar, A.C. Effects of minimum drinking age laws on alcohol use, related behaviors, and traffic crash involvement among American youth 1976–1987. *J Stud Alcohol* 52(5):478–491, 1991.

Organization for Economic Cooperation and Development. *Young Driver Accidents.* Paris, 1975.

Perkins, H.W., and Berkowitz, A.D. Stability and contradiction in college students' drinking following a drinking-age law change. *J Alcohol Drug Educ* 35(1):60–77, 1989.

Preusser, D.F., and Williams, A.F. *Sales of Alcohol to Underage Purchasers in three New York Counties and Washington D.C.* Arlington, VA: Insurance Institute for Highway Safety, in press.

Robertson, L.S. Blood alcohol in fatally injured drivers and the minimum legal drinking age. *J Health Polit Policy Law* 14(4):817–825, 1989.

Robins, L. Study childhood predictors of adult antisocial behaviour: Replications from longitudinal studies. *Psychol Med* 8(4):611–622, 1978.

Rooney, J.F., and Schwartz, S.M. The effects of minimum drinking age laws upon adolescent alcohol use and problems. *Contemp Drug Prob* 6(4):569–583, 1977.

Rosenbluth, J.; Nathan, P.E.; and Lawson, D.M. Environmental influences on drinking by college students in a college pub: Behavioral observation in the natural environment. *Addict Behav* 3(2):117–121, 1978.

Roy, M.B., and Greenblatt, E. *Driving Under the Influence of Liquor: Follow-up Study of Age, Sex, and Simultaneous Offenses.* Commonwealth of Massachusetts, Office of Commissioner of Probation, November 27, 1979.

Saffer, H., and Grossman, M. Beer taxes, the legal drinking age, and youth motor vehicle fatalities. *J Legal Stud* 16(2):351–374, 1987*a*.

Saffer, H., and Grossman, M. Drinking age laws and highway mortality rates: Cause and effect. *Econ Inquiry* 25(3):403–417, 1987*b*.

Schmidt, W., and Kornaczewski, A. The effect of lowering the legal drinking age in Ontario on alcohol-related motor vehicle accidents. In: Israelstam, S., and Lambert, S., eds. *Alcohol, Drugs, and Traffic Safety, Proceedings of the Sixth International Conference on Alcohol, Drugs, and Traffic Safety.* Toronto: Addiction Research Foundation, 1975. pp. 763–770.

Schroeder, J.K., and Meyer, E.D. *Influence of Raising the Legal Minimum Drinking Age in Illinois.* Illinois Department of Transportation, Division of Traffic Safety, December 1983.

Semlitz, L., and Gold, M.S. Adolescent drug abuse. *Psychiatr Clin North Am* 9(3):455–473, 1986.

Shattuck, D., and Whitehead, P.C. *Lowering the Drinking Age in Saskatchewan: The Effect on Collisions Among Young Drivers.* Saskatchewan Department of Health, May 1976.

Smart, R.G. Changes in alcoholic beverage sales after reductions in the legal drinking age. *Am J Drug Alcohol Abuse* 4(1):101–108, 1977.

Smart, R.G., and Finley, J. Changes in drinking age and per capita beer consumption in ten Canadian provinces. *Addict Dis* 2(3):393–402, 1976.

Smart, R.G., and Schmidt, W. Drinking and problems from drinking after a reduction in the minimum drinking age. *Br J Addict* 70(4):347–358, 1975.

Smith, D.I. Effect on non-traffic accident hospital admissions of lowering the drinking age in two Australian States. *Contemp Drug Prob* 13(4):621–639, 1986.

Smith, R.A.; Hingson, R.W.; Morelock, S.; Heeren, T.; Mucatel, M.; Mangione, T.; and Scotch, N. Legislation raising the legal drinking age in Massachusetts from 18 to 20: Effect on 16 and 17 year olds. *J Stud Alcohol* 45(6):534–539, 1984.

Stall, R.; Wiley, J.A.; McKusick, L.; Coates, T.J.; and Ostrow, D. Alcohol and drug use during sexual activity and compliance with safe sex guidelines for AIDS. *Health Educ Q* 13(4):359–371, 1986.

U.S. Department of Health and Human Services. *Seventh Special Report on Alcohol and Health.* DHHS Pub. No. (ADM)90–1656. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1990.

van de Goor, L.A.M.; Knibbe, R.A.; and Drop, M.J. Adolescent drinking behavior: An observational study of the influence of situational factors on adolescent drinking rates. *J Stud Alcohol* 51(6):548–555, 1990.

Vingilis, E., and Smart, R.G. Effects of raising the legal drinking age in Ontario. *Br J Addict* 76(4):415–424, 1981.

Voas, R.B., and Moulden, J. Historical trends in alcohol use and driving by young Americans. In: Wechsler, H., ed. *Minimum-Drinking-Age Laws.* Lexington, MA: D.C. Heath and Co., 1980. pp. 59–72.

Wagenaar, A.C. Effects of an increase in the legal minimum drinking age. *J Public Health Policy* 2(3):206–225, 1981.

Wagenaar, A.C. Aggregate beer and wine consumption: Effects of changes in minimum legal drinking age and a mandatory beverage container deposit law in Michigan. *J Stud Alcohol* 43(5):469–487, 1982*a*.

Wagenaar, A.C. Public policy effects on alcohol consumption in Maine and New Hampshire 1970–1980. *Contemp Drug Prob* 11(1):3–20, 1982*b*.

Wagenaar, A.C. *Alcohol, Young Drivers, and Traffic Accidents: Effects of Minimum Age Laws.* Lexington, MA: D.C. Heath, 1983.

Wagenaar, A.C. Preventing highway crashes by raising the legal minimum age for drinking. *J Safety Res* 17(3):101–109, 1986.

Wagenaar, A.C.; Bracht, N.F.; Boyte, H.C.; Finnegan, J.R.; Forster, J.L.; Heberle, J.A.; Hourigan, M.E.; Jones, R.; Klepp, K.I.; Michels, P.A.; Murray, D.M.; Perry, C.L.; and Wolfson, M. Communities Mobilizing for Change on Alcohol: A Randomized Community Trial. Working paper. 1991.

Wagenaar, A.C.; Finnegan, J.R.; Wolfson, M.; Anstine, P.S.; Williams, C.L.; and Perry, C.L. Youth alcohol access: Where and how adolescents obtain alcoholic beverages. *Public Health Rep*, in press.

Wagenaar, A.C., and Maybee, R.G. The legal minimum drinking age in Texas: Effects of an increase from 18 to 19. *J Safety Res* 17(4):165–178, 1986.

Wagenaar, A.C., and Wolfson, M. Trade-offs between science and practice in the design of a randomized community trial. *Proceedings of the Second International Research Symposium on Experiences with Community Action Projects.* San Diego, CA, January 29-February 2, 1992, in press.

Warren, R.A.; Simpson, H.M.; Page-Valin, L.; and Collard, D. *Point Zero Eight and the Change in Drinking Age: One Step Forward and Two Steps Backward?* Traffic Injury Research Foundation of Canada, 1977.

Whitehead, P.C. *Alcohol and Young Drivers: Impact and Implications of Lowering the Drinking*

Age. Department of National Health and Welfare, Health Protection Branch, Non-medical Use of Drugs Directorate, Research Bureau, Ottawa, 1977.

Whitehead, P.C.; Craig, J.; Langford, N.; MacArthur, C.; Stanton, B.; and Ferrence, R.G. Collision behavior of young drivers: Impact of the change in the age of majority. *J Stud Alcohol* 36(9):1208–1223, 1975.

Wilkinson, J.T. Reducing drunken driving: Which policies are most effective? *South Econ J* 54(2):322–334, 1987.

Williams, T.P., and Lillis, R.P. Changes in alcohol consumption by 18-year-olds following an increase in New York State's purchase age to 19. *J Stud Alcohol* 47(4):290–296, 1986.

Williams, A.F.; Rich, R.F.; Zador, P.L.; and Robertson, L.S. *The Legal Minimum Drinking Age and Fatal Motor Vehicle Crashes.* Washington, DC: Insurance Institute for Highway Safety, 1974.

Williams, A.F.; Zador, P.L.; Harris, S.S.; and Karpf, R.S. *The Effects of Raising the Legal Minimum Drinking Age on Fatal Crash Involvement.* Washington, DC: Insurance Institute for Highway Safety, 1981.

Wolfe, A.C., and Chapman, M.M. *1971 and 1973 ASAP Surveys: Washtenaw County High School Students.* The University of Michigan Highway Safety Research Institute, 1973.

Womble, K. *The Impact of Minimum Drinking Age Laws on Fatal Crash Involvements: An Update of the NHTSA Analyses.* National Highway Traffic Safety Administration Technical Report (DOT-HS-807-349). Springfield, VA: National Technical Information Service, January 1989.

Yamaguchi, K., and Kandel, D.B. Patterns of drug use from adolescence to young adulthood: II. Sequences of progression. *Am J Public Health* 74(7):668–672, 1984.

Zylman, R. When it is legal to drink at 18: What should we expect? *J Traffic Safety Educ* 20(4):9–10, 1973.

Zylman, R. Drinking and driving after it's legal to drink at 18: Is the problem real? *J Alcohol Drug Educ* 20(1):48–52, 1974.

Zylman, R. Collision behavior of young drivers: Comment on the study by Whitehead et al. *J Stud Alcohol* 37(3):393–401, 1976.

Zylman, R. The consequences of lower legal drinking ages on alcohol-related crash involvement of young people. Rebuttal to Dr. Richard Douglass. *J Traffic Safety Educ* 24(2):13–15, 1977.

8

Measurement Issues in Alcohol Survey Data

Jody L. Sindelar, Ph.D.

INTRODUCTION

Economists are increasingly applying economic principles and perspectives to studies of alcohol consumption, alcoholism, alcohol abuse, and alcohol dependence. Economic analysis of alcohol and individuals has been aided by the growing availability of household surveys that include data on alcohol. A variety of measures of alcohol use and abuse can occur across surveys or can be contained in a single survey. The particular alcohol measure used and how it is specified can substantially affect the conclusions drawn from economic analyses. The purposes of this chapter are to identify standard alcohol measures and provide citations to their use, illustrate the importance of the choice of indicator, and identify areas in which related data improvements could be most useful to economists.

This chapter discusses the measurement and use of alcohol indicators from large household survey data sets for social science research, especially economic research. Measures of alcohol use and abuse typically found in such surveys are of two broad types: consumption-based indicators and diagnosis-related indicators. There are a wide variety of measures within each type. The importance of the choice of alcohol indicator is illustrated in this chapter by applying data from the 1988 National Health Interview Survey (NHIS) Alcohol Survey to the issue of the association of income and alcohol use, abuse, and dependence. Also discussed are several types of data that would be useful to economic analyses of alcohol-related issues but are typically not available in existing data sets, for example, data on assets and family members' alcohol use and labor supply. Finally, a few relevant data sets are identified.

The first section of this chapter presents and discusses alternative alcohol indicators. The following section illustrates the importance of a careful choice of alcohol indicators by demonstrating that the indicator chosen may affect the research results and conclusions. The last section identifies other data useful for studying economic issues in alcohol, especially the relation between alcohol use and income.

MEASURING ALCOHOL USE AND ABUSE

There are two broad types of household survey data on alcohol use. The first type

includes data on quantity, frequency, circumstances, and type (wine, beer, and spirits) of alcohol consumed. The second includes measures of alcohol abuse and dependence and problems related to drinking. These categories are delineated for purposes of discussion, but note that a single survey may include data of both types as well as other alcohol-related data (e.g., family drinking patterns). Some overlap may exist between these categories.[1]

Measuring Use of Alcohol—Consumption-Based Indicators

Survey data on alcohol consumption

Household surveys vary in the depth and detail of questions that ask about alcohol consumption. They also differ in the degree of emphasis they give to related issues, such as the respondent's experience with various alcohol-related problems, family drinking history, attitudes toward alcohol use, expectancies about the behavioral effects of drinking, and perceived risks associated with drinking. Surveys have included questions on beverage type (beer, wine, and distilled spirits), maximum amount consumed on any given day or drinking occasion, drinking history (usually including age at first drink), the usual amount drunk per drinking day or occasion, the number of drinking days in the past week and the amount drunk on each day, and the contexts of recent drinking occasions. Different surveys have included various combinations of these questions (National Center for Health Statistics 1973, 1978; Malin et al. 1986; Thornberry et al. 1986; Elainy et al. 1990; Center for Human Resources Research 1987; Caetano 1989).

A commonly used set of questions asks respondents how often they drink each of beer, wine, and spirits and how many drinks they usually consume on a day when they drink each beverage type (National Center for Health Statistics 1978; Malin et al. 1986; Thornberry et al. 1986). In other surveys, respondents are asked about the number of drinks without regard to beverage type (National Center for Health Statistics 1973).

Room (1990) reviewed different questionnaire types and suggested that questions could be asked on volume, frequency, variability, and maximum number of drinks on heavy drinking occasions. Volume can be determined by asking about the usual quantity, the proportion of

[1] *As an alternative to questionnaire data, "apparent consumption" is used to measure the quantity of alcohol consumed at the State level. Apparent per capita consumption is calculated using State alcohol sales tax data divided by the State's relevant population. There are a variety of weaknesses in these data that are well known (see US DHHS 1990) and include biases due to the effects of tourism, nontaxed alcohol (e.g., homemade alcohol, duty-free purchases, and the black market), border crossing, spoilage and spillage, inventories, gifts across State lines, and the distribution of consumption across individuals within States (e.g., imputing consumption to abstainers). The recent controversy over the census and potential biases in the population count may add to the list of potential weaknesses in the denominator (i.e., divide sales by the population). Other "denominator" issues include the lower bound for the age to use (is 14 too young?) and the important issue of the abstainers and the distribution across individuals. Also, if one is interested in ethanol per se, beverages that are taxed the same do not all have the same ethanol content.*

occasions that the individual drank a given level of drinks, and the frequency of drinking at specific levels. Frequency can be determined in a variety of different ways, including asking for a list of occasions in a specified time period. Variability is meant to capture the range of drinking; for example, some individuals will drink very little on most days but will drink heavily on weekends, whereas others will have drinking patterns that are more evenly distributed. Questions on the maximum number of drinks consumed per occasion reflect an interest primarily with those drinkers whose heavy drinking imposes social costs.

The recall period for consumption is often 2 weeks, 4 weeks, or a month, but it can be as long as a year. The shorter timeframes may not represent the typical pattern; however, longer timeframes may give rise to greater recall bias (i.e, one either underestimates because one forgets, or overestimates because one reports "memorable" events which occurred before the period of interest).

Some questions may be included about lifetime drinking patterns, such as, at what age did you have your first drink, or when did you stop drinking? Other surveys, such as the National Health and Nutrition Expenditure Survey (NHANES), use a diary. There may also be questions about why the respondent abstains (if he or she does), with possible answers including "I am an alcoholic" or "It's against my religion."

Quality of data

Analysts who intend to use alcohol consumption data generated by surveys of the kind described here need to be aware of and cautious about a number of quality of data issues that typically affect such surveys. The following remarks are specific to household survey collection of alcohol data.

Underreporting is believed to be a severe problem when gathering alcohol consumption data. Individuals may underreport their consumption so that their behavior appears to conform better to that which is socially desirable. Manning et al. (1989), for example, inflated their NHIS ethanol consumption data by 2.5 times in order to adjust for the potential underreporting. In addition to being potentially large in magnitude, underreporting is likely to be systematically related to characteristics of the individual, such as consumption level, religion, gender, and age. Moderate and average drinkers, for example, might have little reason to underreport their consumption because their consumption level conforms to socially acceptable amounts. Teenagers, however, may overreport in order to exaggerate their drinking prowess or establish their maturity. Issues relating to denial complicate the reporting bias further. Denial is thought to be common among alcohol-dependent individuals. If denial takes the form of denying that one's actual consumption poses a problem, heavy drinkers might not necessarily underreport. However, comparison of aggregate data on sales of alcohol with estimates of consumption based on survey self-reported data suggests that there is underreporting in aggregate (Midanik 1982; Manning et al. 1989). Further research is needed in this area.

The issue of self-reports has received substantial attention (see Babor et al. 1987; Czarnecki et al. 1990; Popham and Schmidt 1981). However, much of the literature on self-reports refers to data collected in clinical settings (for a summary, see Babor et al. 1983). Studies of the validity[2] and reliability of self-reports of consumption in general population surveys also exist (see Williams et al. 1986; Midanik 1982, 1989).[3] For both clinical and survey settings, there is debate on the validity and reliability of self-reports. While most studies have focused on white men, the validity and reliability may vary across ethnic groups and by gender.

Nonresponses on alcohol questions can be expected to be relatively high, because alcohol-related issues are sensitive or private issues for some respondents. In addition, some questions may have even higher nonresponses due, for example, to poor recall for questions referring to lengthy time horizons (e.g., when did you first drink?) and for very specific questions (e.g., what was the largest number of drinks you have had on any one occasion?). Apparently, very little of the vast literature in this area is devoted to the issue of nonresponse. It would be useful to know, for example, how nonresponses vary with the type of question (e.g., how specific or how sensitive an area), the ordering of questions, and the socioeconomic and demographic characteristics of respondents. The latter is especially important because of potential biases introduced in analyses.

Aggregating to total ethanol

Responses to specific survey questions about alcohol consumption are seldom analyzed in their raw form. Instead, they typically are aggregated into indices of overall consumption. Data on the quantity and frequency of beer, wine, and spirits consumed are collapsed, for reasons of parsimony and general ease of use, into a single measure of ethanol consumption per time period, often cast as average daily consumption (Caetano 1989; Hilton 1988*a*; Knupfer 1989; Malin et al. 1986; Williams et al. 1986; Wilsnack et al. 1986). Weights that approximate the average ethanol content of each beverage are used to aggregate to a total ethanol intake.[4]

In some surveys, data from the series of items on consumption of wine, beer, and spirits can be used with little ambiguity to calculate total ethanol consumption.[5] In other cases, there can be some or even

---

[2] Babor et al. (1987) discusses various types of validity with regard to alcohol.

[3] It is hard to find a true "gold standard" against which reported drinking can be compared. In clinical data sets, sometimes information from collaterals is gathered, but these collaterals (e.g., spouses) may also either underreport on purpose or not know the extent of the drinking (Midanik 1989). The threat to respondents of gathering collateral data may be a means of increasing the accuracy of responses.

[4] The weights that are used in the example given later in this chapter are beer = 0.045, wine = 0.129, and spirits = 0.411 (US DHHS 1990).

[5] Even in these cases, however, because the ethanol figures are derived from a series of questions, sometimes the total ethanol consumption that is constructed from these responses is beyond that which is physically reasonable. In addition, because a series of questions is used to formulate the aggregate, there is the risk that data will be missing in at least one of the questions.

substantial complications. For example, questions that ask "how much do you usually drink?" give rise to ambiguity, because not every drinking occasion will conform to the usual. In other cases, the responses may have been recorded as categorical data. For example, when asked how often he or she drank one or two drinks, the respondent may be given the response choices of "nearly every time," "more than half the time," "less than half the time," "once in a while," and "never" (Hilton 1987). In cases like these, assumptions must be made in order to calculate a numerical value for consumption. The accuracy of final estimates depends on the appropriateness of the assumptions made.

Summary measures of the sort just described are useful and relatively easy to apply in estimation. Although convenient, these averages and aggregates may suppress important, relevant information. Instead of the average consumption of ethanol, it may be more important to determine which drinks were consumed (e.g., differences between different beer and wine effects on income or different drink-specific elasticities of demand), whether the drinking occurred in connection with other activities (e.g., drinking and driving or drinking on the job versus after work), and the temporal distribution of drinks (e.g., infrequent binge drinking versus low daily consumption).

The importance of these issues depends on what one is trying to investigate and the policy goals one wishes to address. The temporal distribution, for example, may not be relevant information from a revenue-raising perspective.[6] However, if one is concerned about the social consequences of drinking, it may be important to distinguish between, for example, seven drinks in one night, and one drink on each of seven consecutive nights (see Knupfer 1984). Also, it is possible that drinking patterns may vary over the lifecycle; therefore, the choice of indicator may depend on the age of the population under scrutiny. For example, college students tend to concentrate their drinking more heavily on weekends than do middle-age individuals (Baer 1991).

Thus in some circumstances, instead of using total ethanol consumption, one should retain and use some of the detailed data on type of drink, drinking patterns, beverage preferences, drinking in combination with other activities, etc. The advantages of working with such detailed data must be balanced against the added complexity they bring, both econometrically and in the portrayal of findings. However, these more detailed data may be helpful in understanding issues and developing relevant policy.

Establishing cutpoints for heavy drinkers and abstainers

Cutoff points in the distribution of consumption are sometimes used to focus on only the heavy drinkers, those who pose the greatest social costs. This approach

[6] *From a revenue-raising perspective, the Government may be interested in the distribution as well because a tax increase could affect different types of drinkers differently. One might want to allow for varying price elasticities across groups.*

assumes that there is some threshold effect; that is, alcohol consumption per se is not a problem, but rather heavy drinking is the problem. Making continuous variables into binary (or categorical) data reduces the information available, but it also helps overcome some practical problems such as how to treat influential outliers and inaccurate data on the exact number of drinks. It also helps mitigate some of the problems in dealing with skewed data.[7] Hilton (1988*c*), for example, follows a fairly broadly used convention of defining heavier drinking as one ounce or more of ethanol consumed per day.[8]

Obviously the choice of cutpoint will affect the percentage of the population that is considered to be heavy drinking, and will affect different subgroups differently. A smaller percentage of females, for example, will be found to be heavy drinkers at any cutoff. The appropriate cutpoint could vary with the application. For example, Knupfer (1987) advocates using a relatively high cutpoint so that the indicator will be sensitive to changes in proportions of the heaviest consumers. For other purposes, one may be better off using a cutpoint that incorporates a somewhat larger percentage of the population. Knupfer (1989) suggests using several categories of heavy drinkers, with the heaviest defined as drinking at least eight drinks per occasion three times a week. Ideally, the choice of cutpoint could be analyzed with sensitivity analysis to determine how the results vary with the choice for different age, race, and gender populations under study.

At the opposite end of the drinking spectrum, it has been estimated that about a third of the adult population abstains from alcoholic consumption (Hilton and Clark 1987).[9] Abstention has been defined differently across studies, often as drinking either no alcohol or one drink or less in the last year, or as consuming 12 drinks or less in a lifetime (Williams and DeBakey 1992). Older individuals and females are more likely to abstain than others, ceteris paribus. Abstainers may comprise an analytically interesting subpopulation. For instance, Manning et al. (1992) suggest that not only do heavy drinkers have different elasticities of demand (as often thought), but so do abstainers as compared with moderate drinkers.

Considering the context of alcohol consumption

Alcohol consumption may pose greater problems when combined with other

---

[7] *The distribution of individuals by their consumption is known to be skewed with a mass of observations lying at the zero point (abstainers), and there are some very heavy drinkers at the other tail. Even among abstainers there may be heterogeneity; those who do not currently drink can be divided into former drinkers and those who never drank. As shown below, this designation based on past drinking may be relevant in some contexts.*

[8] *Heavy drinkers can be defined in alternative ways. Hilton (1988a) uses 4 separate definitions: consuming 60 drinks or more per month, having 5 or more drinks per occasion at least once a week, having 8 or more drinks per day at least once a week, and getting drunk at least once a week.*

[9] *This can present difficulties in modeling certain issues; see Manning et al. (1992) for a study on the elasticity of demand that pays specific attention to abstainers.*

activities. Drunk driving, drinking on the job, and drinking while caring for children are prime examples of the use of alcohol in potentially hazardous contexts. In some individual-level data sources, there are questions that elicit information about the context of drinking, that is, alcohol consumption combined with these other activities (e.g., the 1988 NHIS asks about drunk driving). Here the issue of response bias is likely to be of importance, because individuals are likely to underreport their abuse of alcohol (e.g., drinking and driving). Furthermore, the biases in reporting are likely to vary by the characteristics of the respondent. It is possible, however, that the responses themselves, including their biases, could be of interest to the analysts because the responses may reveal information about the respondent, for example, knowledge of risks. Mullahy and Sindelar (1991*b*) found that only a relatively small percentage of those who admit to driving drunk also admit (that is, either are aware of or are willing to admit) that they engage in behavior while drinking that could harm themselves (55 percent of men and 48 percent of women) or harm others (33 percent of men and 27 percent of women).

Alcohol quantity and quality in determining consumption

Economists are becoming increasingly aware that alcohol is heterogeneous with regard to "quality." Quality variation is an issue when using data on drinks because each drink may not conform to the standard in ethanol content. Although the ethanol content may average across drinks, it does not necessarily average across individuals. Some individuals may systematically select drinks with high or low ethanol content even within the beer, wine, and spirits categories (consider wine coolers and Cisco, the new high-ethanol drink, at the opposite ends of the ethanol spectrum among wine beverages). Similar issues arise when using expenditure data. A high expenditure on alcoholic beverages does not always translate into a high level of ethanol consumption. Ignoring these "quality" aspects can be misleading both in estimating the effects of ethanol on outcomes and in estimating elasticities of demand (see Deaton 1988 for more general applications of quality variations). A high-income person may spend more on alcoholic beverages without necessarily consuming more ethanol if, for example, he or she consumes expensive wines or spirits with standard ethanol contents.

Distribution of alcohol consumption across individuals

Perhaps more than other users of alcohol survey data, economists are sensitive to some of the statistical problems that are raised by the lack of normality in the distribution of most alcohol consumption variables. The absence of normality in the consumption data across individuals creates several problems. First, it violates the normality assumptions underlying standard econometric procedures and thus can pose problems in using conventional analysis. Methods for dealing with skewed data include taking the log of the

distribution (making some adjustment for the large number of zeros), transforming the data into categories of drinking levels, and using binary indicators (such as heavy drinkers).

The second and perhaps more complex issue to deal with statistically is that although standard estimation techniques estimate the mean effects of alcohol, the mean effect may mask the heterogeneity of the effects across different drinking levels and patterns. For example, moderate drinking may have different effects from heavy drinking; that is, moderate amounts of alcohol may not be damaging to productivity, whereas heavy consumption might be. Furthermore, the effects of alcohol may vary over the productivity distribution as well. The productivity of an otherwise relatively unproductive person may be nearly unaffected by the development of alcoholism while the corresponding impact could be great for a highly productive individual.

To address the possibility that relationships between alcohol consumption and other variables of interest may differ qualitatively or quantitatively depending on the level of consumption, one could estimate separate regressions of the effects of alcohol using low, moderate, and heavy drinking subpopulations, for example. However, selecting the sample on the basis of these variables can be problematic in estimations. In addition, attention should be paid to the proper delineation of categories so that only those with similar effects would be combined. More sophisticated methods could also be used. For example, one could estimate endogenous switching regressions analyzing heavy drinkers and others separately. However, the data requirements for such techniques may not be met by the available data sources (see the section entitled "Other Data Needs").

Measuring Alcohol Abuse and Dependence—Diagnosis-Based Indicators

Classification of alcohol-related disorders

The second major type of alcohol measures contained in alcohol survey data are measures of alcohol-related problems or disorders. Diagnosing disorders through surveys of the population is a more recent practice than is gathering consumption data through surveys. Diagnoses of disorders used to be confined to a clinical setting. However, without diagnosis of general (as opposed to clinical) populations, prevalence rates in a general population would not be known and analyses based on a random sample of the general population would not be possible. Thus, to make diagnoses for large, population-based samples, clinical standards have been translated into questions and algorithms suitable for survey data. However, with these measures, as with the consumption measures, there are some methodological and statistical issues of concern.

The term "alcoholism" is commonly used to denote an alcohol-related disorder or syndrome. The term often has been used somewhat loosely and is not defined by well-specified criteria. Over time, the term has been replaced in the alcohol field

by a number of others that apply specific criteria for diagnosis. Two of the most common approaches to defining alcohol-related disorders come from the American Psychiatric Association (APA) and the World Health Organization (WHO). The APA's *Diagnostic Statistical Manual (DSM) of Mental Disorders* provides classification of and specific criteria for alcohol dependence and alcohol abuse. (An individual can have either or both disorders, but dependence is considered the more severe, so if the individual meets the criteria for both, he or she is considered to be dependent.) The most recent version of these diagnostic criteria is the DSM-III(R) (APA 1980); a newer DSM-IV version is currently under development. The WHO uses the *International Classification of Diseases* (ICD) for classifying two similar disorders: alcohol dependence and harmful use of alcohol (ICD-10 version). The DSM and the ICD criteria are somewhat different, but the two are converging with their latest forthcoming versions (ICD-10 and DSM-IV). (See US DHHS 1990 for more on this subject).

As there are relatively few data sets combining socioeconomic and demographic data with diagnoses of alcohol-related syndromes in a population-based data set, the social scientist researcher will typically not make a selection among the two classification systems in a single data source. Instead, he or she will have to accept the system contained in available data sets. In the epidemiologic catchment area (ECA) data,[10] for example, the public use version of the data set has variables indicating diagnoses based on the DSM-III system.[11] The ECA uses the diagnostic interview schedule (DIS) to make diagnoses from the battery of self-reported responses to the questionnaire. The DIS is a structured interview that can be administered by lay people[12] yet conforms to the APA's DSM-III criteria for diagnoses. The 1988 NHIS Alcohol Survey, by contrast, asks about a battery of symptoms that were meant to conform to the DSM or ICD classification systems;[13] however, no diagnosis is available on the tape, nor is an algorithm provided to make the diagnosis. However, Grant et al. (1991) calculate prevalence rates using the 1988 NHIS data and discuss an algorithm that can be used to make diagnoses of alcohol abuse and dependence with this data set.

When using clinically based definitions of diagnoses with household survey data, some adjustments must be made.

---

[10] *The ECA is a population-based data set that includes a wide variety of mental health variables, among them indicators of alcohol abuse and dependence. See table 6 for more on this and other data sets.*

[11] *The DSM-III criteria were the immediate predecessors of DSM-III(R).*

[12] *In a structured interview, the responses by the individual are merely recorded; no judgment by the interviewer is used. In contrast, in semistructured interviews, the interviewers must use clinical judgment in response to the respondent's answers to the questions. Lay interviewers thus cannot be used to administer semistructured questionnaires. See Grant and Towle (1990) for more on standardized structured and semistructured diagnostic interviews for alcohol abuse and dependence.*

[13] *In the empirical example given in the following section, the DSM-III(R) criteria are used to formulate diagnoses from these data.*

For example, the DSM-III(R) definition of alcohol abuse and dependence has the requirement that symptoms have existed for a month or longer. However, household surveys typically ask only about the frequency of symptoms rather than their duration. Thus, frequency criteria (e.g., two times per year or more) can be used instead of duration criteria. Grant et al. (1991) use this approach.[14] In the ECA algorithm, in specifying "ever" meeting criteria for diagnosis, it is not required that all the criteria were met at the same time in the individual's life; instead, it is necessary that an individual had ever met all the criteria, even if the symptoms occurred at different times in his or her life.

The timeframe for symptoms in such surveys is often both "in the last year" and "lifetime" (e.g., in the ECA and 1988 NHIS data). The category "in the last year" is obviously useful for its currency, while the "lifetime" or "ever" category may be useful in analyzing lifetime aspects of alcohol-related diagnoses. In addition, because some believe that an alcoholic never is "cured" but rather is in remission or recovering, the lifetime data may be particularly relevant. The lifetime indicator may also be preferred as it is less likely to be codetermined with indicators of current behavior.

Quality of survey data

As with the consumption measures, measures of alcohol abuse and dependence are subject to a number of concerns regarding data quality. Deriving a diagnosis of alcoholism or alcohol abuse on the basis of household survey data is a complicated task that falls to psychiatrists, survey methodologists, and others with expertise in this area. Economists typically do not know the specifics of DSM-III(R)/DSM-IV or ICD-9/ICD-10, or the merits of using DIS or other batteries of questions used for diagnoses.[15] Instead, they would like to be able to use data, for example, on alcohol dependence and abuse, with confidence that they are valid, reliable, specific, and sensitive measures that serve the intended research or policy purposes.

There are, however, potential pitfalls in not understanding the underlying diagnoses, and there may be alternative ways to use the available data. For example, the DSM-III diagnosis of alcohol dependence and abuse uses some symptoms that could vary by location and over time depending on the social and political/regulatory environment. The problem here is that a given individual's diagnosis could vary depending on where and when the diagnosis was made, while it should vary only with the individual's symptoms. Examples of such symptoms include getting arrested by the police, having friends and family suggest that you stop drinking, and having trouble at school or at work due to drinking. In addition, these symptoms may be endogenous to the question under study; for

[14] *They use the criteria of a symptom occurring two or more times per year. The exception to this is for the category "tolerance," in which case only one or more occurrences are necessary.*

[15] *There are several other instruments that have been developed as diagnostic or screening tools. Most of these, however, are not used in large, population-based surveys. See US DHHS 1990 (p. 190) for a listing of these.*

example, estimating educational attainment as a function of diagnosis of alcoholism, which uses the symptom "trouble at school" to make the diagnosis. It has been suggested that the DSM-III be revised in the DSM-IV to make adjustments for the use of symptoms that vary with the social and political setting (see Kosten and Kosten 1986; Rounsaville et al. 1986).

Validity and reliability studies have been conducted for survey data on a variety of diagnoses. Here the "gold standard" is considered to be psychiatrists' reports. It should be noted, however, that there can be substantial disagreement across psychiatrists in the diagnosis of a given individual. The DSM-based diagnoses made by the DIS have been analyzed using the ECA data (Anthony et al. 1985; Helzer et al. 1985; Shrout et al. 1987). Although the validity and reliability of the DIS diagnoses vary by the disorder (depression, anxiety, etc.), alcohol dependence and abuse have been among those found to be the most reliable and valid of the diagnoses. However, these studies have focused on the current (in the last year) diagnosis. The "lifetime" diagnoses have not been scrutinized in the same way; it would be difficult to derive a "gold standard" for these diagnoses.

There is little or no literature on other aspects of the quality of data. Missing responses are a potential problem because some of the questions refer to what would be considered by many to be socially undesirable behavior (e.g., getting the "shakes" after having too much to drink, and drinking and driving). A respondent could either refuse to answer or underreport the extent of this type of behavior. Systematic nonresponses and underreporting could pose problems in analysis. A difficult dilemma is how to treat missing data on individual symptoms when using a complex algorithm for diagnoses (e.g., Grant et al. 1991 do not discuss how missing data are treated). For example, cases that have missing data are often dropped from analyses. However, there may be sufficient information in the remaining responses to classify some of these dropped cases as meeting the diagnostic criteria.

Extensions of the use of diagnostic data

Data from symptom lists are typically aggregated into binary diagnoses of such concepts as alcohol abuse and dependence. An awareness of the procedures and the pitfalls of this aggregation process can be of importance to users of alcohol survey data. Diagnostic data reported from a survey are binary by their very nature and relate only to the tails of the distribution; that is, diagnosis occurs only when the symptoms are severe and numerous enough to indicate a substantial problem with alcohol. The use of diagnostic data is consistent with both the idea that there are threshold effects and also that individuals in the upper tail of the distribution are those who impose the greatest social costs. However, extensions of the use of diagnostic data might include using subclinical symptom levels, thus allowing inquiry into the effects of less severe forms. If, for example, alcohol dependence reduces income, would subclinical manifestations reduce income, but

by a lesser extent? In addition, it might be possible to use the symptom data to measure more severe forms of alcoholism—that is, the greater the number of symptoms, the more severe the case that is diagnosed. However, there is little guidance in the literature as to how to formulate either the subclinical or more severe manifestations.

There may also be some room for innovation in using the data on individual symptoms as opposed to using the diagnosis. Mullahy and Sindelar (1991*b*) used responses to specific questions on drunk driving and perceptions of alcohol-related risks to study drunk driving. Frank and Gertler (1991) used symptom counts instead of diagnoses. Symptom counts may aid in measuring severity and may help to quantify subclinical cases. However, the use of symptom counts in essence assumes that the symptoms have equal weights; this assumption has not been subject to scrutiny and is contrary to the assumptions underlying diagnostic algorithms.

Other diagnosis-related issues of concern include determining the most succinct way to design questionnaires to diagnose alcoholism, thereby allowing questionnaire space for more indepth income and labor supply questions (see the section entitled "Other Data Needs" for more on this). A second issue of concern is identifying the most relevant questions to capture the lifecycle course of diagnoses and symptoms.

CHOICE OF ALCOHOL INDICATOR—AN EMPIRICAL ILLUSTRATION

To illustrate how the choice of alcohol measures can affect conclusions drawn, I present a specific example. The example springs from the debate in the literature as to whether empirically alcohol consumption has a positive or negative effect on income. I use the alcohol/income nexus as the substantive issue and the NHIS data as the empirical basis to demonstrate the importance of the choice. The substantive issue is used only as an example of how the choice of measurement may matter; the empirical analysis is not designed to "solve" the issue but rather to illustrate a point. Resolving the alcohol/income conundrum would take a much more careful, considered, and complicated approach (see Mullahy and Sindelar 1989, 1990, 1991*a*, in press *a*,*b*; and Mullahy, chapter 6). Here the intention is merely to see how much the results can differ according to the indicator used.[16]

Researchers and policymakers have been interested in assessing the "economic costs of alcohol," and more specifically, the "effects of alcohol on income." The phrases "the costs of alcohol" or "the

[16] *Motivation for this approach, the broader interpretation of the coefficients, and the caveats for such interpretations will not be discussed here because we are not trying to estimate the effect of alcohol on income per se. Interpretation of the results as such demands a much more careful approach; one should consider which age categories are being pooled, whether one is measuring the direct or indirect effect, which other alcohol-related variables are included, lifecycle aspects, whether one is including physical and mental comorbidities, etc.*

effects of alcohol" by themselves provide little guidance about which indicator would be best to use.[17] The ambiguity in the specification of "alcohol" as either the level of consumption or as meeting certain diagnostic criteria may be the root of seemingly conflicting results in the literature. Rice et al. (1990) and Mullahy and Sindelar (1989, 1990, 1991*a*, in press *a,b*) used a diagnosis of alcohol abuse and dependence as their alcohol indicator and found negative effects on income.[18] In contrast, Berger and Leigh (1988) and Cook (1991) used consumption of ethanol indicators and found positive effects of alcohol on income. While there are several other differences among these studies (e.g., data source and selection criteria), the difference in the type of alcohol indicator used across these two sets of results may account for some of the disparity of results. This is because diagnoses, by their very nature, capture the tail end of the drinking distribution while per capita consumption indicators may cover the entire drinking distribution. If, for instance, there are nonlinear or specifically inverse U-shaped effects of alcohol consumption on income, analyses based on diagnostic indicators may show only the linear effect, instead of differentiating the positive and negative effects.

Thus, as reflected in the conflicting results in the literature, the choice of indicator can be important and may affect the overall conclusion. I demonstrate this empirically below using the 1988 NHIS data.

Alternative Alcohol Indicators from the NHIS Data

The 1988 NHIS Alcohol Supplement[19] allows comparisons of alternative measures of both alcohol consumption and alcohol abuse and dependence in a single data source,[20] and thus it is suitable for the illustrative task at hand. The advantage of demonstrating the issue using a single data source is that, unlike a comparison across studies, the data source, the sample selection criteria, the control variables, etc. are held constant, and only the alcohol indicators vary.

Table 1 lists and defines[21] a set of alternative measures derived from the 1988 NHIS. The measures include ethanol consumption (per 2-week period), indicators of heavy drinking (using

---

[17] *In some instances, the research question is specific enough to direct the researcher to the use of a particular measure (e.g., what are the effects of taxes on daily consumption) but in other instances, there may be ambiguity. The researcher often has only limited access to data and may have little choice in the selection of alcohol indicators and related variables.*

[18] *Similarly, Berry and Boland (1977) and Harwood et al. (1984) found negative effects, but they used a self-reported determination of alcohol-related problems.*

[19] *The National Health Interview Survey (NHIS) is a national survey conducted yearly by the National Center for Health Statistics. The 1988 survey had an NIAAA-funded Alcohol Supplement which interviewed 43,809 individuals. See table 6 for more on this and other data sets.*

[20] *A single data set may have only a few related measures. The ECA data set, for example, has DSM-based diagnoses, but has no data on consumption. The 1988 NHIS has a variety of variables including consumption by drink, extensive family background on problem drinkers, and symptom data. See table 6 for more on data availability.*

two different cutoff points), DSM-based assessments of alcohol abuse and dependence for both the current year and ever in life (using alternative strictness criteria for symptoms), and symptom counts.

I defined two different types of indicators depending upon whether they were binary indicators or continuous variables. I first looked at binary indicators of the tail of the distribution; these are defined in table 1 (they are indicated with a "D" for dichotomous variables). The binary indicators are based in some cases on alcohol consumption (heavy drinker) and in other cases on diagnoses (dependence and abuse). The proportions of male, female, and all respondents meeting the criteria for each binary variable are displayed in table 2A. As can be seen, the percentage of the population that is assessed as falling into the upper tail of the distribution varies, depending upon how one defines the tail (i.e., by variations in diagnosis and/or consumption of ethanol[22]) and by the relevant time span measured.[23] Obviously more individuals are assessed as being dependent or abusive when using the time span of "ever" as compared to the most recent year.

Table 2B shows the means for each measure in the set of continuous variables. Note, however, that comparisons across the means are not very informative, as the variables are measured in different units. Ethanol consumption per time period is a commonly used measure of alcohol consumption. The measure of alcohol consumption used here was the natural log of amount of ethanol (in ounces) consumed in the last 2 weeks.[24] Also specified are counts of symptoms for either "year" (current) or "ever" (lifetime). Three different specifications for each "year" and "ever" were used corresponding to counts of (1) all the symptoms asked in the NHIS, (2) only those symptoms used in the DSM-based diagnoses, and (3) the nine DSM criteria for dependence. Symptom counts, however, are not commonly used measures.

Variation of Results Across Alcohol Indicators

I used the NHIS data to estimate income regressions using alternative alcohol mea-

[21] *Note that according to the DSM criteria, if an individual meets the criteria for both abuse and dependence, the dependence criteria take precedence and thus the individual is considered to be dependent and not abusive; that is, the categories are made mutually exclusive. For the "year" variables, I displayed the means for anyone who met the criteria, that is, in the not-mutually-exclusive form. When abuse and dependence were entered in the regressions separately, again they were used in the not mutually exclusive specification. However, when they were entered in the same regression, they were made into mutually exclusive categories. For abuse and dependence "ever," I did not make these mutually exclusive on the grounds that at various points in an individual's lifetime, he or she could have had either or both abuse and dependence, and this cannot be determined from the responses.*

[22] *Two types of upper tail individuals can be defined, heavy drinkers and those with drinking problems. These are not completely overlapping categories. For more on this, see Harford et al. (1989), Hilton (1987), and Mäkelä and Simpura (1985).*

[23] *Table 2 also shows how some of these indicators vary by gender and confirm the well-known fact that men are heavier drinkers and are more likely to abuse alcohol than are women regardless of the indicators selected.*

[24] *To avoid the problem of trying to take the log of 0, I used 0.5 in place of 0 for the abstainers.*

TABLE 1

Alternative measures of alcohol

Variable name	Type[a]	Definition
LOGETH	C	Log of total ethanol consumed in 2-week recall; weights of 0.045 ethanol per oz. of beer, 0.129 per oz. of wine, and 0.411 per oz. of spirits
DEPEVER[b]	D	Ever in lifetime satisfied the DSM-III(R) criterion for alcohol dependence
ABUEVER[b]	D	Ever in lifetime satisfied DSM-III(R) criterion of alcohol abuse in the last year
TDEPYR[b]	D	Satisfied criterion for alcohol dependence during the year using the more stringent criterion of counting a symptom only when it occurred at least twice in a year
TABUYR[b]	D	Satisfied criterion of alcohol abuse in the last year counting a symptom if it occurs two or more times
NYSMPTEV1[b]	C	Number of the 41 symptoms available in the NHIS data that were experienced by the respondent in lifetime
NYSMPTEV2[b]	C	Number of the DSM-III(R)-based symptoms ever experienced by the respondent
NYSMPTEV3[b]	C	Number of the nine DSM-III(R) diagnostic criteria satisfied by the respondent ever
FORMERDR	D	The individual is self-proclaimed to be a former drinker
HVDRNK (1 or 2)	D	Drank 10.8 (#1) or 14 (#2) ounces or more of ethanol per 2-week period

[a]D = Dichotomous where 1 = yes, 0 = otherwise; C = Continuous variable.

[b] The algorithm used in formulating the diagnoses corresponds to that used in Grant et al. (1991). However, the "ever" designations vary in two ways: (1) no data on the number of occurrences were available in NHIS, and (2) abuse and dependence are not mutually exclusive—because of the lifetime aspects, we allowed for some individuals to have had abuse and/or dependence at different points in their lifetime.

[c] In a parallel fashion NYSYMPTYR1, NYSYMPTYR2, and NYSYMPTYR3 are defined for symptoms occurring during the year instead of ever in the individual's lifetime.

sures. Males and females were analyzed separately, as there is evidence that they not only exhibit different drinking patterns (Clark and Midanik 1982; Hilton 1988*a*; Malin et al. 1986; Williams et al. 1986; Wilsnack 1986), but estimated income equations for each tend to be significantly different (Mullahy and Sindelar 1991*a*). The log of income was estimated as a function of alternative alcohol indica-

TABLE 2

Means for selected alcohol measures: Males, females, and entire population[a]

Variable	Males	Females	All
A. Dichotomous variables			
HVDRNK1	0.17	0.05	0.10
HVDRNK2	0.13	0.03	0.07
DEPEVER[b]	0.32	0.13	0.21
ABUEVER[b]	0.49	0.19	0.31
TDEPYR[b]	0.10	0.03	0.06
TABUYR[b]	0.12	0.03	0.07
B. Continuous variables			
LOGETH	0.68	–0.07	0.24
NYSMPTEV1	6.26	2.55	4.09
NYSMPTYR1	2.35	1.02	1.57
NYSMPTEV2	4.38	1.77	2.85
NYSMPTYR2	1.56	0.66	1.03
NYSMPTEV3	1.97	0.83	1.30
NYSMPTYR3	0.78	0.34	0.52

[a] *The total number of observations vary somewhat by the number of missing observations, but the range is approximately 41,400 to 42,600 observations. There are approximately 17,200 to 17,700 valid observations for males, and 24,200 to 25,000 for females.*
[b] *These categories are not mutually exclusive; that is, there are individuals in the abuse category who are also in the dependent category. Thus the percentage that are either abusive or dependent is considerably less than the sum of the two. In regressions when they are both included together, the categories are mutually exclusive.*

tors; the coefficients across the alternative alcohol measures are the focus of this example. That is, I estimated a series of regressions in which only the alcohol indicator changes. I compared the coefficients of the alcohol indicators across regressions to see how they differ.

I included a sparse set of exogenous variables in the regression in addition to the alternative alcohol indicators. Included were age, age squared, dummy variables on race, and regional indicators.

Table 3 provides evidence that the conclusions one could reach on the effect of alcohol on income do depend on the measure of alcohol used. Separate regressions were run for each alcohol indicator, although alcohol abuse and alcohol dependence measures were included together in a single regression as well as separately. The coefficients on the alcohol indicators from these separate regressions are displayed in table 3. Table 4 shows the full set of estimated coefficients for only one of the indicators of table 3, the log of ethanol consumed in the last 2 weeks. Parallel regressions were estimated for the other alcohol indicators but are not displayed due to space limitations.

As seen in table 3, the coefficients on the alcohol indicators vary not only in magnitude but also in sign. Although most of the measures of the "tail" of the distribution have negative and significant coefficients, the coefficient on the log of ethanol is positive and significant. Curiously, when both "ever" indicators of alcohol abuse and dependence are included in a single regression, each is significant but each takes on a different sign.

TABLE 3

Ln of income[a] regressions with alternative measures of alcohol use and abuse

	Males		Females	
	Coefficient	t-statistic	Coefficient	t-statistic
LOGETH	0.01	2.47	0.04	6.82
DEPEVER[b]	–0.10	6.56	–0.07	3.90
ABUEVER[b]	0.09	5.11	0.13	5.92
TDEPYR[b]	–0.21	7.54	–0.20	5.01
TABUYR[b]	–0.01	0.34	–0.10	2.43
DEPEVER[c]	–0.12	9.05	–0.09	5.03
ABUEVER[c]	–0.02	1.33	–0.02	1.30
TDEPYR[c]	–0.22	10.28	–0.26	8.41
TABUYR[c]	–0.14	6.99	–0.22	6.96
NYSMPTYR2	–0.02	9.76	–0.003	9.64
NYSMPTEV2	–0.01	11.38	–0.01	6.55
HVDRNK1	–0.04	2.61	–0.06	2.11
HVDRNK2	–0.07	3.65	–0.03	1.05

Note: As noted in the text, this analysis is intended only to illustrate the significance of choosing an appropriate measure of alcohol involvement. The results should not be interpreted as evidence of the effect of alcohol on income.

[a]*Natural logarithm (ln) of income is adjusted by the State Consumer Price Index (CPI).*

[b]*DEPEVER and ABUEVER are included together in the same regression as are TDEPYR and TABUYR.*

[c]*Each of these variables is entered in the regression separately.*

When ABUEV (alcohol abuse ever) is entered separately, its coefficient is never significant.

I explored the ethanol consumption indicator further because the results in this case deviated from the results of the other regressions examined. Specifically, I examined whether or not there were nonlinear effects of ethanol consumption. For this regression, the consumption of alcohol in the past 2 weeks was broken down into the following categories: (1) no current consumption and not a former drinker, (2) no current consumption but a former drinker, (3) 0 to 5.4 ounces of ethanol consumed, (4) 5.4 to 10.8 ounces of ethanol con-

TABLE 4

Sample regressions including all independent variables: Log of ethanol as an example of regressions estimated for all other ethanol measures

	Males		Females	
	B	t	B	t
LOGETH	0.01	2.47	0.04	6.82
AGE	0.04	45.16	0.08	44.40
SQAGE	–0.0009	46.08	–0.009	48.56
WHITE[a]	0.23	5.72	0.25	6.92
BLACK[a]	–0.22	5.13	–0.34	8.66
NORTHEAST[a]	0.03	1.36	–0.04	2.33
MIDWEST[a]	–0.04	2.36	–0.11	6.04
SOUTH[a]	–0.00	0.00	–0.06	3.38
INTERCEPT	7.91	132.74	8.07	144.17

Note: As noted in the text, this analysis is intended only to illustrate the significance of choosing an appropriate measure of alcohol involvement. The results should not be interpreted as evidence of the effect of alcohol on income.

[a] Omitted category is "other" for race and "West" for the regions.

sumed, and (5) 10.8 ounces or more. The omitted category was 0 to 5.4 ounces.

Table 5 shows that all included categories have a negative and significant effect in income regressions. That is, compared to the omitted category of "moderate drinkers," all other categories have negative effects. As stated above, these coefficients do not necessarily portray the "true" relationship as more issues would have to be addressed. Instead they are forwarded to show how alternative specifications of the alcohol variable can affect the conclusions drawn.

Thus, it appears that the choice of alcohol indicator can have an important impact on the conclusions drawn, at least in the example above. To summarize, these results suggest that care must be exercised by the analyst in the choice of alcohol indicator, that the robustness of the results across alternative specifications is not necessarily assured and should be investigated, and that designers of questionnaires should carefully consider the types of data collected in their surveys.

OTHER DATA NEEDS

Empirical analyses of alcohol issues using household survey data suffer a variety of problems due to data constraints. These include, among others, the statistical issues of simultaneity,[25] causation, "the errors in variables problem," and "the omitted variables problem."[26] Although analytical methods to deal with these

[25] *For example, in the alcohol/income issue, does alcoholism cause low income or does low income cause frustration and then alcoholism, or are alcohol and income connected via some alternative route? Longitudinal data might help to solve this conundrum.*

[26] *If, for example, use of illicit drugs is omitted from income regressions, the alcohol indicator could (to the extent that use of drugs and alcohol are positively correlated) capture the effects of both drugs; thus the coefficient could overestimate the extent to which alcohol reduces income. The same could hold for a host of other variables including, for example, other mental and physical health comorbidities, spouse's alcohol consumption, former alcohol consumption, parents' alcohol problems, local labor market conditions, and indices of the cost of living.*

problems are available, data constraints limit their applicability (e.g., in creating good instrumental variables).

Better data in certain areas could improve the usefulness of alcohol-related data sets and open new avenues of research. For the alcohol-income nexus in particular, better data on the following would be useful: (1) individuals' income and labor supply; (2) the labor supply, alcohol use, and other addictive drug use of other family members; (3) variations in alcohol use/abuse over the lifecycle; (4) other addictions and comorbidities of individuals; (5) lifetime aspects of the use of alcohol and other drugs; and (6) geographically varying factors, such as prices of alcohol and local labor market conditions. These are discussed in more detail below. No data set meets all of the requirements for data described below. Some of the publicly available, population-based data sets are listed (with brief comments) in table 6.

TABLE 5

Nonlinear alcohol effects

	Males		Females	
	B	t	B	t
ETHANOL-0	–0.22	11.01	–0.24	15.55
FORMERDR	–0.22	11.83	–0.13	7.37
ETH(5.4-10.8)	–0.08	3.90	–0.10	3.82
ETH(10.8+)	–0.14	7.74	–0.06	2.13

Notes: The omitted category is ethanol consumption in the range of 0 to 5.4 ounces, or a "moderate" drinker. The negative coefficients on the included categories mean that the effects of alcohol are nonlinear; that is, moderate consumption is "beneficial" relative to the other categories.

The number of observations for each category, in order starting with FORMERDR, are for males: 4,498; 10,701; 6,549; 1,271; and 1,165; and for females: 3,559; 3,253; 5,473; 1,993; and 3,037.

As noted in the text, this analysis is intended only to illustrate the significance of choosing an appropriate measure of alcohol involvement. The results should not be interpreted as evidence of the effect of alcohol on income.

Income, Assets, and Labor Supply

From an economist's vantage point, most alcohol (and more generally health-related) household surveys have not collected wage, income, and labor supply data that are sufficiently detailed, reliable, and comprehensive. This is particularly true with regard to the alcohol/income relationship, but it is also problematic in many other areas of interest to economists, such as in estimating demand elasticities. The data on income in health-related surveys are often weak in the following areas: (1) large numbers of nonresponses on income questions; (2) little or no data on hourly earnings or hours worked; (3) little or no data on nonwage income, assets, and wealth; (4) use of family income instead of person-specific data; and (5) reliance on categorical income data.[27]

Although family income data are typically collected in alcohol-related surveys,

[27] *Categorical data pose several problems: (1) continuous data are needed for many applications, (2) the categories are so small at the lower end of the income distribution (e.g., $1,000 increments) that they result in an unnecessarily large number of nonresponses in some cases (due probably to individuals' uncertainty as to exactly which category is correct), and (3) the upper-income cutoff is too low. There are alternative ways to deal with these data weaknesses but collecting better data is a preferable solution.*

TABLE 6

Some data sets that combine alcohol data with socioeconomic and demographic data in a population-based survey

1. Epidemiologic catchment area (ECA)

a.	In sum	Approximately 20,000 observations; 5 data sites; 2 to 3 waves (first 2 waves publically available); 1980–85; primarily to assess mental disorders (see Reiger et al. 1984, Robins et al. 1981)
b.	Alcohol	Diagnosis by DIS of DSM-III alcohol abuse and dependence; no consumption data
c.	Other drugs	Yes
d.	Lifecycle	Age of onset of alcohol abuse and dependence
e.	Family	Interviews individual only; little data on family
f.	Income	Individual and family; categorical data; maximum category $50,000+ (unpublished data, $120,000+)
g.	Employment, labor supply	Limited
h.	Environment	No (can get census tracts for matching)

2. National Health Interview Survey (NHIS), 1988 Alcohol Survey

a.	In sum	National sample; more than 40,000 observations; conducted yearly; various years (e.g., 1983, 1985, and 1988) have alcohol questions; primary focus on health conditions (see US DHHS 1989)
b.	Alcohol	Number of drinks by beer, wine, and spirits; drinking pattern; heavy drinking of family; symptoms for diagnosis
c.	Other drugs	No (smoking from Occupational Health Supplement)
d.	Lifecycle	Some; lived with problem drinker when young; first drank
e.	Family	All family members interviewed for core data; one adult for Alcohol Survey; data on problem drinking relatives
f.	Income	Family income only; categorical data; maximum $50,000+
g.	Labor supply	Limited
h.	Environment	No (by special request to NCHS may get data matched)

TABLE 6 *(continued)*

Some data sets that combine alcohol data with socioeconomic and demographic data in a population-based survey

3. National Household Survey on Drug Abuse, 1988

a.	In sum	National sample of individuals 12 years old and above; more than 8,880 observations; yearly; 1988 through current publically available; larger sample size starting 1991; primarily to assess illicit drug and alcohol abuse (see NIDA 1988)
b.	Alcohol	Alcohol; drinking patterns; symptoms
c.	Other drugs	Better coverage than most surveys
d.	Lifecycle	Age at first drink
e.	Family	Little
f.	Income	Personal earnings; categorical; $50,000+ maximum
g.	Employment, labor supply	Employment; some on chief wage earner also
h.	Environment	No

4. National Longitudinal Survey—Youth (NLS—Y)

a.	In sum	Longitudinal survey of nearly 13,000 individuals aged 14 to 21 in 1971; oversample minorities and poor; special military sample (see Center for Human Resources Research 1987)
b.	Alcohol	Consumption last 30 days; 1982–83 and 1988; 1988 have data on family members drinking; no symptoms; no diagnosis
c.	Other drugs	Illicit drugs
d.	Lifecycle	Longitudinal
e.	Family	Father drinking problem
f.	Income	Better coverage than most surveys
g.	Employment, labor supply	Better coverage than most surveys
h.	Environment	Geographic matching is possible

such data are almost never suitable for measuring productivity or wage rates paid. In some cases, data on hours worked can be used in conjunction with income data to obtain a proxy for hourly wages, which in turn proxies for productivity.[28] More frequently, however, hours data are also missing, or if available, they cannot be matched to the income measure (e.g., with family income). Data on hours worked (and on unemployment) can be important information independent of the calculation of an hourly wage rate. Labor supply and unemployment can be examined separately as outcomes of interest and can be used to interpret income data; for example, are individuals working long hours and thus earning higher total income in order to compensate for a low hourly rate due to alcoholism (i.e., the substitution effect versus the income effect)? Alcohol's most damaging effects may occur through its effects on unemployment (see Mullahy and Sindelar 1992 for some evidence on this).[29] That is, heavy alcohol use may be related to greater periods of unemployment, but not necessarily to lower earnings during periods when the respondent is employed. Good data on spells of unemployment are rarely available in health-related surveys.

The suggestion here is that survey designers should pay more attention to designing wage rate, income, labor supply, and unemployment questions, perhaps learning from other surveys more attuned to these areas, and to devote more survey space to gathering comprehensive measures. Methods could be developed to improve the response rate and accuracy, for example, asking individuals to look at their adjusted gross income on their income tax returns prior to the survey.

Family Members

For many research applications, information on members of the respondent's family is likely to be relevant (even for respondents who are not married and who live alone).[30] There are many reasons for this. For instance, income data are often collected on a family basis, and they are often difficult to interpret without knowing what portion of the income was earned by the respondent. More generally, the behavior of family members is interlinked and the behavior of one member cannot be viewed in isolation.[31] This may apply to decisions to drink, to drink and drive, to drink excessively, as well as to labor/leisure choices. In the case of the alcohol/income nexus, if the alcoholism of

---

[28] Productivity may be the endpoint of most interest. There is the parallel but even more difficult to measure issue of productivity of time at home.

[29] As unemployment refers to those who seek but do not find work, the most damaging effect may be due to lack of employment, but I use the term unemployment for ease of exposition.

[30] This of course applies to other types of data as well. It is especially clear that this is true for alcoholism, as it is known that alcoholism runs in families. Attributing an individual's lower income to his or her own alcoholism can be misleading if other members also suffer. An individual without an alcohol problem may be married to an individual with a problem, which would not be revealed in the data without including such questions. Mullahy and Sindelar (1992) used the NHIS data to analyze how parents' as well as own alcoholism affects the probability of being employed, unemployed, or out of the labor market.

the spouse is an omitted piece of information yet the alcoholism reduces the income of the nonalcoholic, the researcher may underestimate the impact of alcoholism.[32] Thus data on family members' income, labor supply, alcohol consumption, and alcoholism as well as other data would be useful information. One possibility is for household surveys to interview all adults in the family as well as to ask questions about the other family members. In the right circumstances (e.g., private interviews), the latter could be used for cross-validation of individuals' answers.

The Lifecycle

Alcoholism and alcohol consumption appear to vary over the lifecycle, yet the importance of lifecycle effects and cohort effects have not been clearly disentangled. Lifecycle aspects of drinking behavior deserve further attention; for example, to what extent do previous drinking patterns and/or the previous occurrence of alcoholism have current effects? Early onset of alcoholism may reduce educational attainment and increase marital instability, among other things, so that even for alcoholics in remission, earnings would be lower than for a comparable person who never had alcoholism (Mullahy and Sindelar 1990, in press *b*). Also, there may be cumulative effects of alcohol consumption and alcoholism on outcomes of interest. That is, the effects may not be felt contemporaneously but rather with a lag. Those who consume little or no alcohol currently may still suffer the effects of past consumption, for example, as suggested by the former drinker indicator for females.

Data limitations have restricted analyses of the lifecycle aspect of alcohol consumption and the magnitude of persistent effects. Thus, to improve current knowledge, lifecycle data should be gathered in these areas. Longitudinal data (like the National Longitudinal Survey of Youth, for example; see table 6) would be desirable from an analytical perspective. However, longitudinal surveys are expensive and time-consuming, and the expense of interviewing often reduces the number of respondents surveyed. Alternatives to new longitudinal surveys would be to include more questions on retrospective behavior in cross-sectional data. This too has weaknesses, such as recall bias, but efforts to increase the reliability of data may lessen some of these problems. Another possibility would be to add alcohol questions, including retrospective and family data, to ongoing

[31] *The 1988 NHIS Alcohol Survey asks questions about the other family members' drinking patterns but only interviews one adult per family so not all relevant data are gathered.*

[32] *If one family member abuses alcohol and consequently earns less per hour and works fewer hours, this could have several impacts on other family members. The other family members may work more hours in the market in order to compensate for the lower income (and possibly to avoid a less pleasant situation at home). On the other hand, the effect could be just the opposite—the nonalcoholic may want to spend more hours at home to compensate for the alcoholic's lower productivity at home. In either case, there is likely to be a cross-family member effect. This effect could vary systematically by socioeconomic and demographic factors.*

longitudinal collection efforts. A third possibility would be to reinterview individuals once or periodically (as in the NHANES followup) asking about the intervening time period as well as gathering current alcohol-related data.

Comorbidities and Use of Other Drugs

Individuals who use alcohol frequently also tend to use other drugs, and alcoholics often have other mental health comorbidities. With the omission of data on use of other drugs and on comorbidities, the coefficient on alcoholism in regressions could pick up the effects of these, to the extent that they are positively correlated. Thus data on the use of other drugs and also on other comorbidities could be important information.[33] Furthermore, in some cases, such information on family members would be important. If illicit drugs and alcohol are either substitutes or complements, then data on cross prices also become relevant.

Geographically Varying Factors

Many factors in the economic, regulatory, and even physical environment can influence both the use of alcohol and the effect of alcohol on income and other outcomes. Data on these could be useful as control variables and also as instrumental variables to address the simultaneity issue. There are a variety of alcohol and other related regulations that could also affect consumption of alcohol and its effects. Desired data would include prices of and excise taxes on beer, wine, and spirits; drunk driving laws; and limitations on sales of liquor (e.g., hours and location restrictions). Even the physical environment, such as the weather, could affect both consumption and its effect on income. Nonalcohol-related data, such as indicators of the local labor market, can be important in studying income. Such data could either be supplied directly on the tape, or identifiers (e.g., city, county, or State) could be provided to allow researchers to match the data themselves.

CONCLUSION

A variety of indicators of alcohol use (patterns of drinking, ethanol consumption, drinking context, etc.) and alcohol abuse and dependence are available in several large population-based household survey data sets. The choice of which indicator to use should be made with care and knowledge of the characteristics of the indicators; selections should be analyzed for robustness of the conclusions drawn. The empirical exercise using the Alcohol Supplement to the 1988 NHIS shows that the choice of alcohol indicator and how it is used can affect fundamental conclusions drawn. Comparisons across studies should take into account alternative indicators used.

Although there are several large household data sets that combine alcohol indicators with socioeconomic and demographic data, all have some weaknesses

[33] *The ECA Survey has data on other mental health comorbidities. The National Institute on Drug Abuse National Household Drug Abuse Survey has ample data on the use of other drugs. The NHIS has detailed data on physical health.*

that could constrain analysis. Some of the vexing problems of causation, simultaneity, and inference could be solved, and all would be helped, by gathering additional relevant data. Areas of need in economic research (especially regarding the alcohol/income nexus) include improved data on income, labor supply, family members' behavior and characteristics, use of other drugs, indicators of other comorbidities, and relevant geographically varying factors.

ACKNOWLEDGMENTS

I thank Greg Bloss and Michael Hilton for their suggestions on the revision of this paper; John Mullahy for his discussions in reference to our joint work; Stephanie O'Malley for helping to formulate the algorithms for diagnosis of alcohol abuse and dependence; and Zhiwie Ma for programming. This chapter was prepared with support from National Institute on Alcohol Abuse and Alcoholism grant AA-08394.

REFERENCES

American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders.* Third Edition. Washington, DC: American Psychiatric Association, 1980.

Anthony, J.; Folstein, M.; Romanski, A.; VonKorff, M.; Nestadt, G.; Chahal, R.; Merchant, A.; Brown, H.; Shapiro, S.; Kramer, M.; and Gruenberg, E. Comparisons of the DIS and a standardized psychiatric diagnosis. *Arch Gen Psychiatry* 42:667–675, 1985.

Babor, T.F.; Hesselbrock, M.; Hesselbrock, V.; Meyer, R.; and Workman, K. Never believe an alcoholic? On the validity of self-report measures of alcohol dependence and related constructs. *Int J Addict* 18(5):593–609, 1983.

Babor, T.F.; Stephens, R.S.; and Marlatt, G.A. Verbal report methods in clinical research on alcoholism: Response bias and its minimization. *J Stud Alcohol* 48(5):410–424, 1987.

Baer, J.S. Implications for early intervention from a biopsychosocial perspective on addiction. *Behav Change* 8(2):51–59, 1991.

Berger, M., and Leigh, J.P. The effect of alcohol use on wages. *Appl Econ* 20:1343–1351, 1988.

Berry, R., and Boland, J. *The Economic Costs of Alcohol Abuse.* New York: Free Press, 1977.

Caetano, R. Drinking patterns and alcohol p .-lems in a national sample of U.S. Hispanics. In: Spiegler, D.; Tate, D.; Aitken, S.; and Christian, C., eds. *Alcohol Use Among U.S. Ethnic Minorities.* National Institute on Alcohol Abuse and Alcoholism Monograph 18. DHHS Pub. No. (ADM)89–1435. Rockville, MD: National Institute on Alcohol Abuse and Alcoholism, 1989. pp. 147–162.

Center for Human Resources Research. *NLS Handbook 1987: The National Longitudinal Surveys of Labor Market Experience.* Columbus, OH: The Ohio State University, 1987.

Clark, W., and Midanik, L. Alcohol use and alcohol problems among U.S. adults: Results of the 1979 survey. In: *Alcohol Consumption and Related Problems.* Alcohol and Health Monograph 1. DHHS Pub. No. (ADM)82-1190. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1982. pp. 3-52.

Cook, P.J. The social costs of drinking. Proceedings, Expert Meeting on Negative Social Consequences of Alcohol Use. Oslo: Norwegian Ministry of Health and Social Affairs, 1991.

Czarnecki, D.M.; Russel, M.; Cooper, M.L.; and Salter, D. Five-year reliability of self-reported alcohol consumption. *J Stud Alcohol* 51(1): 68–76, 1990.

Deaton, A. Quantity, quality and spatial variation of price. *Am Econ Rev* 78(2):418–430, 1988.

Eliany, M.; Giesbrecht, N.; Nelson, M.; Wellman, B.; and Wortley, S., eds. *National Alcohol and Other Drugs Survey: Highlight Report.* Ottawa: Health and Welfare Canada, 1990.

Frank, R., and Gertler, P. Assessment of the measurement for estimating the effect of mental distress on income. *J Hum Resources* 26(1):154–164, 1991.

Grant, B.; Harford, T.; Chou, P.; Pichering, R.; Dawson, D.; Stinson, F.; and Noble, J. Epidemiologic Bulletin No. 27, prevalence of DSM-III-R alcohol abuse and dependence: U.S. 1988. *Alcohol Health Res World* 15(1):91–96, 1991.

Grant, B., and Towle, L. Standardized diagnostic interviews for alcohol research. *Alcohol Health Res World* 14(4):340–347, 1990.

Harford, T.; Grant, B.; and Hasin, D. The relationship between ethanol intake and DSM-III alcohol use disorders: A cross-perspective analysis. *J Subst Abuse* 1:231–252, 1989.

Harwood, H.J.; Napolitana, D.M.; Kristiansen, P.L.; and Collins, J.J. *Economic Costs to Society of Alcohol and Drug Abuse and Mental Illness: 1980.* Research Triangle Park, NC: Research Triangle Institute, 1984.

Helzer, J.E.; Stoltzmon, R.; Farmer, A.; Brockington, I.; Plesons, D.; Singerman, B.; and Works, J. Comparing the DIS with a DIS/DSM-III-based physician reevaluation. In: Eaton, W.W., and Kessler, L.G., eds. *Epidemiologic Field Methods in Psychiatry: The NIMH Epidemiologic Catchment Area Program.* New York: Academic Press, 1985. pp. 285–308.

Hilton, M.E. Demographic characteristics and the frequency of heavy drinking as predictors of self-reported drinking problems. *Br J Addict* 82:913–925, 1987.

Hilton, M.E. Demographic distribution of drinking patterns in 1984. *Drug Alcohol Depend* 22(1):37–47, 1988*a*.

Hilton, M.E. Trends in drinking problems and attitudes in the United States: 1979-1984. *Br J Addict* 83(12):1421–1427, 1988*b*.

Hilton, M.E. Trends in U.S. drinking patterns: Further evidence from the past 20 years. *Br J Addict* 83:269–278, 1988*c*.

Hilton, M.E., and Clark, W.B. Changes in American drinking patterns and problems, 1967-1984. *J Stud Alcohol* 48:515–522, 1987.

Knupfer, G. The risks of drunkenness. A comparison of frequent intoxication indices and of population sub-groups as to problem risks. *Br J Addict* 79:185–196, 1984.

Knupfer, G. New directions for survey research in the study of alcoholic beverage consumption. *Br J Addict* 82:583–585, 1987.

Knupfer, G. The prevalence of various social groups of eight different drinking patterns, from abstaining to frequent drunkenness: Analysis of 10 U.S. surveys combined. *Br J Addict* 84:1305–1318, 1989.

Kosten, T., and Kosten, J. Criteria for diagnosis. In: Miller, N.S., ed. *A Handbook of Drug and Alcohol Addiction.* New York: Marcel Dekker, 1986.

Mäkelä, K., and Simpura, J. Experiences related to drinking as a function of annual alcohol intake by sex and age. *J Drug Alcohol Depend* 15:389–404, 1985.

Malin, H.; Wilson, R.; Williams, G.; and Aitken, S. 1983 Alcohol/health practices supplement. *Alcohol Health Res World* 10:48–50, 1986.

Manning, W.; Keeler, E.; Newhouse, J.; Sloss, E.; and Wasserman, J. The taxes of sin: Do smokers pay their way? *JAMA* 261(11):1604–1609, 1989.

Manning, W.; Blumberg, L.; and Moulton, L. "The Demand for Alcohol: The Differential Response to Price." Presented at the Health Economics Meeting, Johns Hopkins University, Baltimore, 1992.

Midanik, L.T. The validity of self-reported alcohol consumption and alcohol problems: A literature review. *Br J Addict* 77:357–382, 1982.

Midanik, L.T. Perspectives on the validity of self-reported alcohol use. *Br J Addict* 84:1419–1423, 1989.

Mullahy, J., and Sindelar, J. Lifecycle effects of alcoholism on education, earnings, and occupation. *Inquiry* 26:272–282, 1989.

Mullahy, J., and Sindelar, J. An ounce of prevention: Productive remedies for alcoholism. *J-Policy Anal Manage* 9:249–253, 1990.

Mullahy, J., and Sindelar, J. Gender differences in the labor market effects of alcoholism. *Am Econ Rev* 81(2):161–165, 1991*a*.

Mullahy, J., and Sindelar, J. Do drinkers know when to say when? An empirical analysis of drunk driving. Submitted, 1991*b*.

Mullahy, J., and Sindelar, J. Employment, unemployment, and alcohol. Unpublished manuscript, 1992.

Mullahy, J., and Sindelar, J. Alcoholism, work and income over the life cycle. *J Labor Econ*, in press *a*.

Mullahy, J., and Sindelar, J. Direct and indirect effects of alcoholism on human capital. *Milbank Q*, in press *b*.

National Center for Health Statistics. *Plan and Operation of the Health and Nutrition Examination Survey, United States, 1971–1973, Vital and Health Statistics.* Series 1, No. 10b. DHEW Pub. No. (HSM)73–1310. Washington, DC: Health Services and Mental Health Administration, 1973.

National Center for Health Statistics. *Plan and Operation of the HANES I Augmentation Survey of Adults 25–74 Years, United States, 1974-1975, Vital and Health Statistics.* Series 1, No. 14. DHEW Pub. No. (PHS)78–1314. Washington, DC: Public Health Service, 1978.

National Institute on Drug Abuse. *National Household Survey on Drug Abuse: Main Findings.* DHHS Pub. No. (ADM)88–1586. Rockville, MD: National Institute on Drug Abuse, 1988.

Popham, R.E., and Schmidt, W. Words and deeds: The validity of self-report data on alcohol consumption. *J Stud Alcohol* 42(3):355–358, 1981.

Reiger, D.A.; Myers, J.; Kramer, M.; Robins, L.; Blazer, D.; Hough, R.; Eaton, W.; and Locke, B. The NIMH epidemiologic catchment area (ECA) program: Historical context, major objectives, and study population characteristics. *Arch Gen Psychiatry* 41:934, 1984.

Rice, D.P.; Kelman, S.; Miller, L.; and Donmeyer, S. *The Economic Costs of Alcohol and Drug Abuse and Mental Illness: 1985.* DHHS Pub. No. (ADM)90–164, Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1990.

Robins, L.N.; Helzer, J.; Weissman, M.; Orvaschel, H.; Gruenberg, E.; Burke, J.; and Reiger, D. National Institute of Mental Health diagnostic interview schedule: Its history, characteristics, and validity. *Arch Gen Psychiatry* 38:381, 1981.

Room, R. Measuring alcohol consumption in the U.S.: Methods and rationales. In: Kozlowski, L., et al., eds. *Research Advances in Alcohol and Drug Problems.* New York: Plenum Press, 1990. pp. 39–80.

Rounsaville, B.J.; Spitzer, R.L.; and Williams, J.B.W. Proposed changes in DSM-III substance use disorders: Description and rationale. *Am J Psychiatry* 143(4):463–468, 1986.

Shrout, P.E.; Spitzer, P.; and Fleiss, J. Quantification of agreement in psychiatric diagnosis revisited. *Arch Gen Psychiatry* 44:172–177, 1987.

Thornberry, O.T.; Wilson, R.W.; and Golden, P. *Health Promotion and Disease Prevention, Provisional Data from the National Health Interview Survey: United States, January-June 1985. Advance Data from Vital and Health Statistics.* No. 119. DHHS Pub. No. (PHS)86–1250. Hyattsville, MD: National Center for Health Statistics, 1986.

U.S. Department of Health and Human Services. *Vital and Health Statistics: Design and Estimation for the National Health Interview Survey, 1985-94.*

Series 2, No. 110. DHHS Pub. No. (PHS)89-1384. Hyattsville, MD: U.S. Department of Health and Human Services, 1989.

U.S. Department of Health and Human Services. *Seventh Special Report to the U.S. Congress on Alcohol and Health.* DHHS Pub. No. (ADM)90–1656. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1990.

Williams, G.D., and DeBakey, S.F. Changes in levels of alcohol consumption: United States, 1983–1988. *Br J Addict* 87(4):643–648, 1992.

Williams, G.D.; Dufour, M.; and Bertolucci, D Drinking levels, knowledge, and associated char acteristics: 1985 NHIS findings. *Public Healt Rep* 101:593–598, 1986.

Wilsnack, S.; Wilsnack, R.; and Klassen, A. Epidemiological research on women's drinking. *In: Women and Alcohol: Health Related Issues.* National Institute on Alcohol Abuse and Alcoholism Research Monograph 16. DHHS Pub. No. (ADM)86-1139. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1986.

☆ U.S. GOVERNMENT PRINTING OFFICE : 1993 O - 360-196 QL II

-5777-

LIBRARY
NIH
Amazing Research.
Amazing Help.

http://nihlibrary.nih.gov

10 Center Drive
Bethesda, MD 20892-1150
301-496-1080

WERT
BOOKBINDING
Grantville, Pa.
JULY - SEPT. 1996
We're Quality Bound

NIH LIBRARY
3 1496 01022 8156